Thèse

présentée en vue de
l'obtention du titre de

Docteur
de
l'Université Paul Sabatier

Spécialité : Sciences Cognitives

Comment déchiffrer le code impulsionnel de la Vision ?
Étude du flux parallèle, asynchrone et épars
dans le traitement visuel ultra-rapide.

---

par

Laurent Perrinet

Soutenue le 7 Février 2003 devant la commission d'examen composée de :

MM.	M.	Imbert	Président
	J.	Hérault	Rapporteur
	Y.	Burnod	Rapporteur
	S.	Thorpe	Co-directeur de thèse
	M.	Samuelides	Directeur de thèse

Thèse

présentée en vue de
l'obtention du titre de

Docteur
de
l'Université Paul Sabatier

Spécialité : Sciences Cognitives

Comment déchiffrer le code impulsionnel de la Vision ?
Étude du flux parallèle, asynchrone et épars
dans le traitement visuel ultra-rapide.

---

par

Laurent Perrinet

Soutenue le 7 Février 2003 devant la commission d'examen composée de :

MM.	M.	Imbert	Président
	J.	Hérault	Rapporteur
	Y.	Burnod	Rapporteur
	S.	Thorpe	Co-directeur de thèse
	M.	Samuelides	Directeur de thèse

à Paul

Remerciements

> Josué ben Pera'hia et Nitaï d'Arbele reçurent d'eux la tradition. Josué ben Pera'hia disait : "Fais-toi un maître, acquiers-toi un compagnon et juge tout homme favorablement".
>
> *Pirke Avot, Traité des Pères*, chap. I, paragraphe 6, Michna 6

Je tiens tout d'abord à remercier Manuel Samuelides, professeur de Mathématiques à Supaéro et chercheur à l'Onéra, pour avoir été mon directeur de thèse. Grâce à ton enthousiasme et ta rigueur scientifique, tu m'as fait découvrir le sens et la beauté de la recherche, merci ! Je te remercie aussi particulièrement pour les gestes qui m'ont guidé dans les moments difficiles. Je remercie aussi vivement Simon Thorpe et toute l'équipe du Cerco qui m'ont accueilli, choyé et guidé tout au long de ce travail. Merci à Rufin pour nos échanges, et merci à Arno pour les multiples discussions et l'accueil à San Diego ! Je souhaite sincèrement remercier Michel Imbert, Jeanny Hérault et Yves Burnod pour avoir accepté de participer à mon jury et aussi pour l'intérêt qu'ils ont manifesté lors de la lecture et le soutenance de ce travail.

Pour m'avoir accueilli pendant ces années au sein du DTIM (Département de Traitement de l'Information et modélisation), laboratoire de l'Onéra, centre de Toulouse, je tiens à remercier René Jacquart alors responsable du département et Daniel Kalfon, coordinateur de notre équipe. Merci à Josette Brial, Noëlle Desblancs et Christiane Payrau pour leur dynamisme et leurs encouragements, ainsi qu'à tous les stagiaires. Merci à Manu pour le déplacement depuis Marseille !

Je pense aussi à tous ceux que j'ai rencontré durant la thèse lors de rencontres imprévues, de conférences ou écoles d'été ; en particulier à Jörg. Muito obrigado à Luis, Matteus, Susana, Yuval, Peter, Anna, Valeria, Sander, Dominique, la famille Jones, Erik, Jeremy, Trisha, Charris et tous les autres !
Un grand merci à ma famille et à mes proches pour leur soutien et leur amitié.

...et surtout, à Cécile.

Table des matières

Résumé **xiii**

Abstract **xvii**

Introduction **1**

1 Comment casser le code de la vision ? **3**

1.1 La vision, un objet d'étude scientifque 3

1.1.1 La vision comme prototype d'un processus cognitif . . 3

1.1.2 Modèles computationnels de la vision 9

1.1.3 L'inspiration biologique en traitement de l'image : vers l'ingénierie neuromorphique 13

1.2 Vers une vision dynamique 16

1.2.1 Quel est le rôle du temps dans le calcul neural ? 17

1.2.2 Quels modèles de codage impulsionnel rapide ? 20

1.2.3 Vers des modèles fonctionnels de la vision 23

1.3 Organisation de la thèse 26

1.3.1 Méthodologie scientifique 26

1.3.2 Plan de l'étude 28

2 Neurophysiologie du système visuel **31**

2.1 La vision comme un système dynamique 31

2.1.1 Quel est le but de la vision ? 31

2.1.2 La vision et son corrélât neuronal 35

2.1.3 Flux d'information dans l'architecture du SNC 40

2.2 Anatomie du système visuel 45
2.2.1 La rétine : de l'information lumineuse au codage impulsionnel 45
2.2.2 Propagation de l'image vers les aires visuelles 51
2.2.3 Représentations temporelles de l'information visuelle . 54
2.3 Contraintes temporelles sur le code visuel 60
2.3.1 Catégorisation Ultra-Rapide 60
2.3.2 Conséquences : une première représentation de la scène visuelle par une vague asynchrone de potentiels d'actions 63
2.3.3 Vers un modèle dynamique de la vision 66

3 Modèles de codage rapide 71
3.1 Modèles de codage temporel 72
3.1.1 Modélisation du mécanisme impulsionnel 72
3.1.2 Codage impulsionnel et adaptation synaptique 86
3.1.3 Codage impulsionnel par la latence de la première impulsion 98
3.2 Modèle dynamique de la rétine 104
3.2.1 Codage multi-échelle de l'information rétinienne 104
3.2.2 Résultats : un code impulsionnel compact 114
3.2.3 Optimisation statistique du modèle dynamique rétinien par les statistiques des images naturelles 119
3.3 Codage par Rang 129
3.3.1 Inspiration biologique du codage par rang 129
3.3.2 Étude théorique du codage par rang 135
3.3.3 Applications 142

4 Vers un code dynamique épars 149
4.1 Pour un modèle d'interactions latérales 149
4.1.1 Densité du codage et représentation sur-complète . . . 149
4.1.2 Codage par poursuite 154
4.1.3 Code impulsionnel épars 159
4.1.4 Application au codage rétinien 163

4.2 Modèle neuronal de représentation corticale sur-complète dans l'aire visuelle primaire (V1) 167
4.2.1 Le modèle de Hubel et Wiesel 168
4.2.2 Modèles d'adaptation 174
4.2.3 Conséquences en traitement de l'image 180
4.3 Systèmes visuels dynamiques et applications 184
4.3.1 Heuristiques attentives et génération de cartes d'attention 185
4.3.2 Code impulsionnel multicouche : vers des représentations corticales 188
4.3.3 Applications à la robotique et à la vision active 193

Conclusion **198**

Abréviations et acronymes **201**

Publications **203**

Bibliographie **224**

Table des figures

1.1.1 Frontières de la vision. 6
1.1.2 Approches sur la nature des calculs dans les processus cognitifs. 15

2.2.1 Organisation axiale simplifiée de la rétine. 46
2.2.2 Les Chemins de la vision. 55
2.3.1 Catégorisation animal/non-animal : exemples d'images utilisées par [Thorpe et al., 1996] 61
2.3.2 Activité différentielle entre les potentiels évoqués (EEG) des cibles et des distracteurs. 62

3.1.1 Modèle impulsionnel de Hodgkin-Huxley. 74
3.1.2 Modèle de neurone impulsionnel, l'Intègre-et-Tire à Fuite. . . 76
3.1.3 Modèle intègre-et-tir à réponse impulsionnelle. 80
3.1.4 Modèle de neurone impulsionnel. 81
3.1.5 Synapses. 85
3.1.6 Facilitation et dépression. 87
3.1.7 Plasticité à modulation temporelle relative : expérience. . . . 89
3.1.8 Plasticité à modulation temporelle relative : résultats. 90
3.1.9 Modèle génératif de la STDP. 92
3.1.10 Détection non-supervisé de cohérence et STDP. 96
3.1.11 Détection de la cohérence par STDP. 97
3.1.12 Luminance et luminosité. 102
3.2.1 Architecture du modèle dynamique de rétine. 106
3.2.2 Filtres linéaires de la rétine. 107

3.2.3 Reconstruction progressive de l'image par les coefficients de la transformée. 112
3.2.4 Mesure de la convergence de la reconstruction progressive de l'image par les coefficients de la transformée. 114
3.2.5 Quantification par leur rang de la valeur absolue des coefficients de la transformée. 118
3.2.6 Mesure de la convergence de la reconstruction progressive de l'image par les coefficients de la transformée. 120
3.2.7 Décorrélation spatiale. 121
3.2.8 Harmonisation des histogrammes par la statistique des images naturelles. 123
3.2.9 Distance subjective entre des images. 126
3.2.10 Quantification de la valeur absolue des coefficients de la transformée par leur rang. 127
3.2.11 Mesure de la convergence de la reconstruction progressive de l'image par les coefficients de la transformée optimisée. . . . 128
3.3.1 Modèle de neurone détecteur d'asynchronie. 133
3.3.2 Principe du codage d'une valeur par son rang. 134
3.3.3 Etude statistique de l'activité d'un neurone à codage par rang. 139
3.3.4 Estimation de la densité de probabilité du temps de tir d'un neurone à codage par rang. 140
3.3.5 Sélectivité Ultra-rapide à l'orientation dans V1. 143
3.3.6 Recherche de concordance. 145
3.3.7 Architecture de SPIKENET. 146

4.1.1 Instabilité de la représentation par ondelette. 151
4.1.2 Principe du codage impulsionnel par poursuite gourmande. . 160
4.1.3 Codage impulsionnel par poursuite α-gourmande. 162
4.1.4 Comparaison ondelettes / poursuite gourmande sur une grille dyadique. 165
4.2.1 Champs récepteurs des neurones de V1 chez le chat. 169
4.2.2 Reconstruction de l'image par une poursuite gourmande soit dans la rétine soit dans V1. 171

4.2.3 Code impulsionnel dans V1. 172
4.2.4 Introduction de connaissance a priori. 174
4.2.5 Auto-organisation spatiale des polarités dans la rétine. . . . 176
4.2.6 Émergence de filtres sélectifs à l'orientation. 179
4.2.7 Émergence d'une carte topologique de filtres sélectifs à l'orientation. 181
4.2.8 Compression par le code impulsionnel épars de l'image *Lena* 183
4.2.9 Détection de contours. 184
4.3.1 Heuristiques de propagation. 187
4.3.2 Carte de saillance dynamique. 189
4.3.3 Système olfacto-visuel Nose. 194
4.3.4 Détection de balises par un robot. 197

Résumé

Les modèles de codage de la Vision, c'est à dire les algorithmes qui extraient des informations pertinentes de l'information visuelle brute produite par des capteurs biologiques (l'œil) ou artificiels (une caméra), se basent aujourd'hui le plus souvent sur des techniques héritées du calculateur de von Neumann et ne répliquent ni l'architecture, ni les performances des systèmes biologiques. En nous fondant sur des expériences neurophysiologiques mettant en évidence le *traitement ultra-rapide de l'information visuelle* [Thorpe et al., 1996], nous allons définir les bases d'un code parallèle et asynchrone de la vision. Celui-ci transforme l'information visuelle analogique organisée spatialement (l'image) en une vague d'événements élémentaires codant grâce à leur succession dans le temps l'information de façon progressive et hiérarchique.

Afin d'explorer les mécanismes neuronaux fondamentaux qui implantent le traitement ultra-rapide, nous allons baser dans un premier temps les mécanismes de ce code sur les processus biologiques à l'échelle du neurone et de la synapse. Cette méthode permettra de définir un modèle génératif de plasticité synaptique [Perrinet et Samuelides, 2000] dont la formalisation et les résultats sont compatibles avec les observations neurophysiologiques [Perrinet et al., 2001; Delorme et al., 2001b]. En particulier, un tel modèle peut induire, grâce à des principes informationnels, l'extraction non-supervisée de l'information cohérente dans le flux des impulsions neuronales [Perrinet et Samuelides, 2002a].

Le codage par la latence de la première décharge peut alors être appliqué à un modèle de la rétine qui permet de définir un algorithme complet de codage et décodage impulsionnel dans le nerf optique [Van Rullen et Thorpe, 2001].

Celui-ci se base sur une architecture multiéchelle similaire à une transformée en ondelettes qui peut exploiter les régularités de ses coefficients sur les images naturelles pour les quantifier par le rang d'arrivée des impulsions neuronales. Nous allons alors étendre cette démarche grâce aux résultats d'Atick [1992] en utilisant une approche *écologique* pour le code visuel impulsionnel. Celui-ci se base sur une optimisation de l'information transmise par les impulsions [Perrinet et Samuelides, 2002b] grâce aux *statistiques des images naturelles*. Cette méthode améliore en théorie et en pratique la régularité des coefficients sur l'ensemble des images naturelles, et donc l'efficacité du code impulsionnel, et définit en outre une mesure de la distance entre des images plus proche de la disparité subjective.

Ce modèle de code visuel permet alors d'envisager la construction d'un modèle impulsionnel multicouche adapté pour la vision et basé simplement sur l'ordre d'activation des neurones. Ce code, par le rang des décharges, est basé sur une architecture hiérarchique et "en avant" qui se distingue, outre sa simplicité, par la richesse des résultats mathématiques et de par ses performances computationnelles. En particulier, l'utilisation d'outils statistiques étendus à ces modèles dynamiques permet de définir et d'optimiser les paramètres macroscopiques des modèles basés sur ce code. On pourra alors implanter des solutions complètes de traitement visuel, à l'image du logiciel SPIKENET [Delorme et al., 1999a] et dont les performances se rapprochent des systèmes biologiques.

Toutefois, les conditions sur l'architecture du modèle de transmission dans le nerf optique impliquent des limites qui ne permettent pas son extension à d'autres aires visuelles. Nous allons répondre aux besoins d'un modèle efficace de la Vision en fondant une théorie de *représentation impulsionnelle surcomplète* de l'image. Cette formalisation conduit alors à une stratégie de *code impulsionnel épars* [Perrinet et Samuelides, 2002c] en définissant de façon rigoureuse des interactions latérales. Cette stratégie peut alors être étendue à un modèle de *colonne corticale adaptative*, grâce à une règle d'apprentissage par renforcement, qui montre des comportements comme l'émergence de filtres corticaux, similaires à la biologie et aux résultats d'autres modèles computationnels, comme ceux de Olshausen et Field [1998]. De plus, ce pa-

radigme s'adapte particulièrement à la construction d'une carte de saillance modélisant de façon dynamique les zones d'attention dans l'image [Perrinet et al., 2002]. Tout en s'intégrant dans le cadre de l'*ingénierie neuromorphique*, ces techniques font émerger de nouveaux outils pour le traitement de l'image dont nous montrons des applications à la détection de contours, à la compression dynamique d'images et enfin à la reconnaissance d'objets appliquée à la robotique.

Mots clés

Vision, traitement visuel ultra-rapide, calcul parallèle et asynchrone, codage neuronal impulsionnel, plasticité hebbienne à modulation temporelle relative (STDP), transformée par ondelettes, statistique des images naturelles, représentation sur-complète, poursuite de concordance, ingénierie neuromorphique, compression dynamique d'images, carte attentive de saillance, détection d'objets, vision active et robotique.

Abstract

How to decipher vision's spiking code ? Study of the parallel, asynchronous and sparse flow in the ultra-rapid visual processing.

Models of visual coding, i.e. the algorithms which extract the relevant perception from the rough spatial and luminous sensation produced by biological (the eye) or artificial (a camera) sensors, are today generally based on techniques inherited from the von Neumann computer and retort neither architecture, nor the performances of the biological systems. Basing us on neurophysiological experiments highlighting the *ultra-rapid processing of visual information* [Thorpe et al., 1996], we will define the bases of a parallel and asynchronous code of the vision. This model transforms the analogical visual information (the image) into a wave of elementary events coding the information thanks to their succession in time in a progressive and hierarchical way.

In order to explore the fundamental neuronal mechanisms which establish this high-speed treatment, we will at first base the mechanisms of this code on the biological processes on the scale of the neurone and synapse. This method will make it possible to define a generative model of synaptic plasticity [Perrinet et Samuelides, 2000] which formalisation and results are compatible with the neurophysiological observations [Perrinet et al., 2001; Delorme et al., 2001b]. In particular, such a model can induce, thanks to informational principles, the non-supervised extraction of coherent information in the flow

of the neuronal impulses [Perrinet et Samuelides, 2002a].
Coding by the latency of the first spike can then be applied to a model of the retina which makes it possible to define a complete spike coding and decoding algorithm in the optic nerve [Van Rullen et Thorpe, 2001]. It is based on multi-scale architecture similar to a wavelet transform and exploits the regularities of its coefficients over the set of natural images to quantize their values by the rank of the spikes. We will then optimize this method according to the results of Atick [1992] by using an ecological approach for the visual spike code. This code is based on optimisation of the information transmitted by the impulses [Perrinet et Samuelides, 2002b] thanks to the *statistics of natural images*. This method improves the regularity of the coefficients on all the natural images, both theoretically and practically, and therefore the effectiveness of the pulse code, and defines in addition a measure of the distance between images closer to subjective disparity.
This visual code model makes it possible to envisage the construction of a multi-layer spiking model adapted to the vision and simply based on the order of activation of the neurones. This code, by the rank of the discharges, is based on a hierarchical feed-forward architecture which is distinguished, in addition to its simplicity, by its mathematical and computational performances. In particular, the use of statistical tools extended to these dynamic models makes it possible to define and optimise the macroscopic parameters of the models based on this code. One will be able then to establish complete solutions of visual processing, as the SPIKENET [Delorme et al., 1999a] software, which compare to biological systems.
However, the conditions on the architecture of this model of transmission in the optic nerve imply limits which do not allow its extension to other visual areas. We will meet the needs for an effective model of Vision by defining a theory of *over-complete event representation* of the image. This formalization leads then to a strategy of a *sparse spike code* [Perrinet et Samuelides, 2002c] by defining in a rigorous way lateral interactions between neurones from a same area. Thanks to a reinforcement learning rule, this strategy can then be extended to a model of an *adaptive cortical column* which shows behaviors like the emergence of cortical filters similar to biology and to the results of

other computational models, like those of Olshausen et Field [1998]. Moreover, this paradigm adapts particularly to the construction of a saliency map modeling in a dynamic way the areas of attention in the image [Perrinet et al., 2002]. Within the framework of *neuromorphic engineering*, these techniques allow emergence of new tools for image processing for which we show some applications to edge detection, dynamic image compression and object recognition applied in particular to robotics.

Keywords

Vision, ultra-rapid visual processing, parallel and asynchronous computing, neuronal spike coding, spike time dependent plasticity (STDP), wavelet transform, natural images statistics, over-complete representation, matching pursuit, neuromorphic engineering, dynamic image compression, saliency map, object recognition, active vision and robotics.

Introduction

La Vision est une porte pour découvrir la cognition et le vivant. C'est à ce titre, mais aussi pour comprendre l'importance de notre compréhension des processus de pensée, que nous l'étudierons et que nous essaierons de bouleverser quelques paradigmes établis sur les processus visuels. En effet, à l'image de nombreux phénomènes liés au vivant, on ne peut réduire la Vision à des lois physiques et on doit faire appel à des méthodologies décrivant son fonctionnement depuis les structures et régularités que l'ont peut observer. Nous tenterons de compléter ces méthodes analytiques et inductives par une approche qui décrit la Vision comme un ensemble de processus émergents de l'interaction avec le monde physique, grâce à des règles et mécanismes simples et interconnectées.

À ce titre, nous étudierons la Vision dans le champ des Sciences cognitives par une approche connexionniste et nous essaierons d'étudier comment les primates et plus particulièrement l'Homme —cet animal qui est resté un chasseur pendant 99% de son évolution— ont intégré dans leur Système Nerveux Central (SNC) des processus permettant d'intéragir avec le monde environnant. Or, si l'on arrive à simuler précisément le comportement individuel des cellules qui constituent la plus grande part du SNC, les *neurones*, la simulation d'un système entier, comme le SNC le plus simple du monde vivant, comportant exactement 531 neurones et dont on connait précisément la topologie, se révèle être un échec.

En effet, il manque à notre connaissance le *code* des processus neuronaux, c'est-à-dire le "mode d'emploi" qui régit leurs comportements. Si ce code existe, il n'est sûrement pas unique dans le monde vivant, mais nous nous attacherons à en décrire des aspects généraux en le décrivant comme un algo-

rithme, *i.e.* une suite d'instructions de calculs, que nous intégrerons dans son implantation physique. En particulier, les phénomènes de traitement visuel ultra-rapide [Thorpe et al., 1996] peuvent se révéler essentiels dans notre compréhension du code visuel. À l'image d'une faille de sécurité exploitée dans une démarche d'ingénierie inverse par un informaticien, nous allons essayer d'extraire des règles générales de fonctionnement en explorant les limites du code visuel.

Chapitre 1

Comment casser le code de la vision ?

> Trouver [...], c'est vaincre une inhibition et une tradition. On ne peut pas avancer si on n'est pas subversif.
>
> *Laurent Schwartz*, in *Le Monde*, 1997

1.1 La vision, un objet d'étude scientifque

La vision va être le centre de cette thèse comme un objet d'étude dans la champ très large des Sciences Cognitives. C'est en la définissant sous la double face de son implantation biologique et de sa modélisation computationnelle que nous allons pouvoir soulever une problématique introduisant une compréhension dynamique et écologique du code de la vision.

1.1.1 La vision comme prototype d'un processus cognitif

La vision est souvent accusée d'être un sens "dictatorial" tant par son importance dans l'Évolution et dans la physiologie du vivant que par l'importante place qu'elle prend dans les études neuroscientifiques. Mais c'est avant

tout une sensation majeure pour découvrir le monde et nous l'étudierons comme un modèle pour la perception et la cognition en général.

Notre vision de la Vision

En effet, en participant de façon active aux processus vitaux et cognitifs, la vision est un modèle en Sciences Cognitives pour comprendre les phénomènes perceptifs et le fonctionnement de la connaissance. Elle permet de construire avec les autres modalités sensorielles une perception stable à partir de l'univers des sensations. Cette stabilité s'exprime notamment par la constance de la perception personnelle, qu'elle soit temporelle — ne voyais-je pas de la même façon quand j'étais enfant ? — ou à plus court terme, avec par exemple la constance de la prception des couleurs malgré des changements de luminosité. Elle existe aussi de façon inter-personnelle, de telle sorte que des personnes ayant une aberration stéréoscopique (qui ne voient donc pas le relief perçu en fusionnant les images des yeux gauche et droit) ne pourront souvent se rendre compte de leur différence perceptive qu'après de nombreuses années en confrontant leur perception par des tests spécialisés. On comprend alors l'importance des lois de la *Gestalt* (de l'allemand *modèle* ou *motif*), c'est-à-dire -en simplifiant- des lois regroupant des percepts en unités indivisibles (un disque, un visage) et qui semblent émerger des fonctions supérieures de la vision. Par analogie avec le langage, on peut imaginer que ces processus utilisent une "grammaire" spécifique au monde des images perçues. La vision serait selon cette définition un processus inférant le monde depuis son image par une suite hiérarchique de règles logiques.
Toutefois, certaines images présentent des illusions optiques qui montrent que cette inférence produite par la vision n'est ni simple ni parfaite. Ces limites se révèlent alors par des distorsions ou des hallucinations qui sont difficiles voire impossibles à éliminer malgré le caractère illogique de leur perception (voir Fig. 1.1.1-(A)). Ce contre-exemple montre alors que la vision est plutôt un *processus inductif*, "devinant" depuis l'information visuelle une interprétation nécessaire aux processus vitaux, mais qui peut-être incomplète ou fausse. La vision comme objet d'étude révèle aussi qu'elle n'est pas unique et qu'elle

s'inscrit dans un contexte. Par exemple, l'ensemble de taches de la figure Fig. 1.1.1-(B) ne semble pas présenter une organisation particulière si l'on ne sait pas à quel contexte il appartient. Par contre, on ne peut pas ne plus voir l'objet qu'elle représente quand on l'a reconnu. La nature de la perception pose alors des questions —qu'est ce que nous percevons ? est-ce un reflet exact du monde ? mon voisin ressent-il la même chose ? et la mouche en volant ?— qui montrent que la vision est un objet d'étude ardu qui reflète notre relative ignorance des processus cognitifs.

L'illusion de l'homuncule

La sensation de l'image mentale et d'une façon plus large des processus mentaux qui sont reliés à cette perception, comme revoir et manipuler mentalement des objets mémorisés ou se balader mentalement dans un endroit connu conduit souvent à penser que la vision, et plus particulièrement la vision humaine, serait le produit d'une représentation interne par l'*homuncule*, ce "petit homme dans notre tête" qui jugerait, analyserait puis traduirait nos sensations. Si une silhouette de cet homuncule existe de fait dans le cortex préfrontal sous la forme d'une carte sensorielle représentant les différentes parties de notre corps (et dont l'aire varie avec l'importance sensitive relative) et qui relie cette partie du cerveau aux muscles correspondants, ce concept est illusoire. En effet, cette vision de notre cognition et de la vision a été largement critiquée pour les contradictions qu'elle entraîne (comme par exemple sa récursivité théorique : l'homuncule a-t-il lui-même un homuncule ?), et par son caractère simpliste. De façon plus pratique, elle est en contradiction avec les observations neurophysiologiques et des troubles liés à des lésions de zones précises du cerveau, par exemple, montrent que des facettes de la perception peuvent être perdues alors que la globalité de la sensation est souvent préservée. La vision et les processus cognitifs que nous étudions sont donc largement intermodaux et n'ont pas de localisation centrale.

L'image perçue par la vision n'est donc pas reconstruite sur un "écran" intérieur, et —paradoxalement— si elle n'est nulle part, où est-elle ? Dans

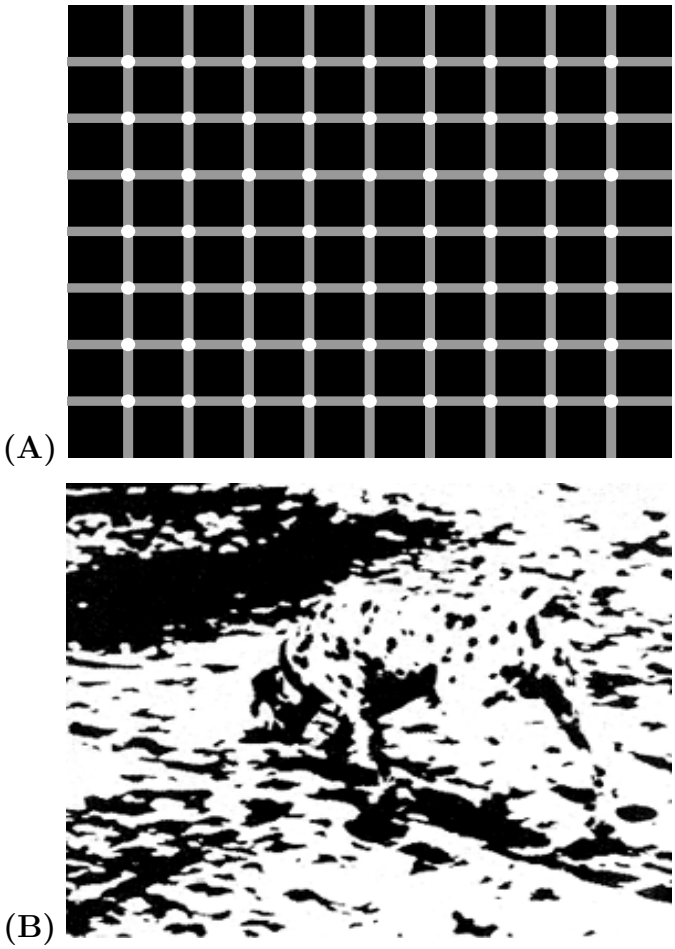

FIG. 1.1.1 – *Frontières de la vision.* Les "illusions optiques" sont un bon exemple d'une confrontation entre l'image et la perception. Ces distorsions, en exploitant des configurations éloignées des images naturelles, peuvent prendre plusieurs formes. **(A)** Les illusions *physiologiques* sont inhérentes à la nature du système visuel et particulièrement aux premières étapes du traitement : si l'on fixe un des points blancs, des points noirs semblent apparaître et clignoter à l'emplacement des points blancs distaux. Cette perception change au gré des saccades oculaires en attirant le regard, mais sont impossible à "rattrapper". **(B)** Les illusions *cognitives* participent à des processus plus complexes. Cette célèbre image de R.C. James présente une scène avec un très fort contraste. On a du mal à percevoir des objets si l'on est novice, alors que si l'on a déjà vu cette image, même quelques années auparavant, on reconnaît parfaitement un chien dalmatien de trois-quart arrière sur un fond ombragé de feuilles.

ce sens, on peut alors s'intéresser à la vision comme le lien entre le monde lumineux et nos actions. Cet aspect de la vision diverge de celle que nous expérimentons par l'imagination en nous remémorant par exemple une image car, par exemple, une part de la vision participant à l'action n'est alors pas forcément consciente. Pour preuve, des images présentées très rapidement en vision périphérique peuvent être catégorisées avec succès alors que les sujets reportent souvent ne pas avoir eu conscience de l'image (alors qu'ils l'ont *perçue*) et malgré souvent leur résistance à donner une réponse qui leur semble totalement hasardeuse [Thorpe et al., 2001b]. En dissociant ainsi la vision de l'image consciente que nous percevons, l'image mentale (ou *eidétique*), nous nous acquittons aussi d'une part de la subjectivité qui entoure sa définition. Cette simplification nous permet d'explorer comment nous pouvons étudier la vision en la restreignant aux processus cognitifs qui permettent de relier l'information lumineuse à des comportements.
Ainsi, dans le cadre d'une expérience neuropsychologique, Bach-y Rita [1983] a stimulé le dos d'un patient aveugle suivant la transcription tactile (par des palpeurs) d'une image grossière captée par une caméra montée sur un trépied : peut-on alors créer une image mentale par d'autres modalités sensorielles, une "rétine tactile" ? Il a alors montré que le sujet ne percevait pas ces images qu'on lui présentait. Par contre, quand découragé, celui-ci prit la caméra en main, alors l'interaction de l'action avec l'image lui a donné une conscience du lien entre le monde réel et l'image : la vision a été créée. On peut alors faire une analogie avec la *salle chinoise* de Searle. Cette expérience de pensée imagine un travailleur scrupuleux enfermé dans une salle et auquel on transmet des feuilles comportant des signes qu'il peut traduire grâce à un dictionnaire approprié. Si les signes sont des calligrammes chinois et que la traduction est bien correcte, on ne pourra toujours pas dire que la personne *sait* traduire le chinois. De la même façon, lorsque nous voyons, il est nécessaire d'agir sur le monde lumineux pour que les processus participant à cette traduction prennent un sens [O'Regan et Noë, 2001].

La vision dans le Système Nerveux Central

Aussi, pour étudier *où* et *comment* ont lieu les processus biologiques participant à la vision, il est alors nécessaire d'étudier les chemins conduisant de l'organe sensitif, *l'œil* jusqu'aux lieux de l'action comme par exemple les muscles du bras qui vont orienter un mouvement. Plus particulièrement, ces chemins font partie du *Système Nerveux Central* (SNC), c'est-à-dire l'ensemble des cellules neurales* regroupées par fibres suivant les nerfs, le long de l'épine dorsale et dans le cerveau. La position classique considère alors le SNC comme un système transformant hiérarchiquement l'information lumineuse depuis l'œil en une activité dans le système visuel, *l'information visuelle* et qui est progressivement traduite suivant le contexte pour être transmise par exemple au système moteur.
Comme cela s'est répété dans l'histoire de la compréhension de la perception, cette "vision" est largement inspirée par la technologie actuelle de calcul. Et par analogie avec les calculateurs modernes, on imagine souvent que la vision consiste en différents processus séquentiels de calcul —à l'image des algorithmes de traitement de l'image— qui vont suivre une suite ordonnée et hiérarchique de processus avant de produire une réponse, e.g. motrice. En se basant sur ce schéma, on peut alors se demander quel est alors le rôle des différentes parties du SNC, mais aussi comment et par quels principes celles-ci se sont organisées au cours de l'Évolution.
Cette définition sous-entend aussi que les réseaux de neurones formant le SNC pourraient se réduire à des processus calculatoires logiques transformant l'entrée visuelle en une réponse comportementale. Toutefois, les neurones ne sont pas seulement des calculateurs et présentent des dissimilarités trop importantes —asynchronie, plasticité— pour être comparés à des opérateurs de logique classique. En effet, si certains de leurs comportements sont des calculs (comme l'intégration d'entrées, ou le seuillage logique de l'activité), leur nature biologique montre, aux nombreux niveaux de leur étude, que leurs comportements sont beaucoup plus riches. Par conséquent, cette richesse pose

*Dans la suite du texte, nous utiliserons, comme en anglais, le terme *neural* pour indiquer les éléments relevant de l'ensemble du SNC, par opposition à *neuronal* que nous restreindrons à ce qui relève du neurone et de ses sous-composants.

la question de la définition et l'implantation biologiques des processus calculatoires inhérents au SNC et qui sont impliqués dans les processus visuels.

1.1.2 Modèles computationnels de la vision

L'histoire de notre connaissance de la Vision montre le lien étroit avec l'histoire des techniques et des modèles des processus cognitifs et visuels.

Vision et Intelligence Artificielle

Tout d'abord, il a fallu que des doutes soient émis sur la validité d'une approche purement symbolique pour qu'apparaissent des solutions imitant les processus émergents observés dans la biologie. Historiquement, les premiers modèles de neurones formels, apparus à l'époque du premier calculateur ENIAC grâce à McCullough et Pitts [1943], sont des portes logiques binaires, ou *switches* généralisant l'entrée à une somme linéaire pondérée. La presse élogieuse y a vu la concrétisation de nombreux rêves de Science-Fiction, mais leurs premières applications semblent seulement apparaître avec le Perceptron de Rosenblatt [1960] dans la même période que la découverte des champs récepteurs du cortex primaire par Hubel et Wiesel [1959]. Toutefois, ces modèles abstraits ne sont alors pas (ou peu) appliqués à la vision. L'approche de Marr [1982] est alors fondatrice, car elle propose une métathéorie en trois étapes pour construire un modèle de vision active. Selon Marr, un modèle du système visuel doit tout d'abord comporter une théorie *computationnelle*, c'est-à-dire définissant le but et quantifiant la pertinence des calculs et représentations possibles. Cette théorie définit donc un *code*, dans le sens informatique d'un algorithme mais aussi dans l'aspect indissociable de la définition de la représentation de l'information qui est manipulée. La seconde étape définit les entrées puis les sorties désirées, et par une approche algorithmique la transformée et la représentation implantant de la meilleure manière ces processus visuels. Enfin, le système doit pouvoir être interprété en terme de composants physiques, et d'une réalisation matérielle. La vision est alors réduite à un processus computationnel. En citant l'auteur,

> Vision is a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information.
>
> *Marr et Nishihara*

Toutefois, cette approche diverge de ce que l'on observe dans la nature, puisque d'une part les implantations pratiques se limitent aux technologies existantes (algèbre linéaire dans le traitement de l'image, calculateurs synchrones) et car d'autre part, elle débute par la définition de son but fonctionnel. Elle aura toutefois dans l'histoire des systèmes de vision artificielle une influence créative et elle a notamment permis de faire émerger par des processus d'optimisation des systèmes se rapprochant de la biologie.
Avec ce cadre épistémologique, Marr a alors défini dans le système visuel les processus suivant quatre niveaux hiérarchiques. Le premier est constitué par l'image brute à partir de laquelle est constituée une *esquisse primitive*. En particulier, le système de détection de contours proposé par Marr [1980] montre que cette approche permet d'obtenir des résultats copiant des aspects du traitement rétinien. Ensuite, cette ébauche permet de construire une représentation en $2^1/_2$D, *i.e.* entre une représentation dessinée ou en volumes, des primitives des objets (contours, texture orientée des surfaces) qui permettent à un troisième niveau de reconnaître ces formes primitives, ou *géons* [Biederman, 1985]. Cette représentation permettrait enfin d'aboutir à un modèle tridimensionnel cohérent de la scène que nous la percevons subjectivement. Ce modèle hiérarchique a beaucoup influencé les systèmes de vision effectivement implantés et a eu en pratique un grand succès.

La renaissance du connexionnisme

L'étude des processus cognitifs par ces modèles révéla un paradoxe toujours vivant : les tâches complexes (jouer aux échecs, calculer des logarithmes) sont plus facilement simulées par un calculateur que des tâches relevant du sens commun, comme dire un texte ou rechercher un visage dans une foule. Ce paradoxe semble intimement lié à la nature de l'approche, le calculateur relevant d'une approche de *haut en bas*, c'est-à-dire en analysant un problème

pour le découper en sous-problèmes plus simples, alors que les processus cognitifs semblent plutôt émerger du *bas en haut*, c'est-à-dire par l'interconnexion de processus élémentaires : c'est l'approche *connexionniste*.
En parallèle avec les progrès de notre connaissance de la vision, les progrès de la technologie du calcul ont permis de recourir à des simulations de plus grande ampleur et de permettre une renaissance du connexionnisme dans les années 1980. L'algorithme de rétropropagation du gradient ou les modèles à mémoires associatives ont alors incité à la création de telles applications qui émergent de processus élémentaires simples et laissé entrevoir la puissance potentielle de tels systèmes. Des applications, comme l'application de conversion de texte à parole *NeTalk*, présentent alors des analogies dans leur performances avec les systèmes biologiques jusque dans leurs imperfections, certaines réalisations de NeTalk souffrant même de "dyslexie".
Le modèle de mémoire associative réalisé par Hopfield [1982] est de ce point de vue exemplaire. Dans ce système, un simple réseau de neurones binaires asynchrones avec des poids symétriques exhibe un comportement de mémoire "adressable" par le contenu. En effet, une fois apprises les connexions entre les neurones, les activités pouvaient se réarranger après la présentation d'un stimulus tronqué ou bruité pour converger vers le stimulus complet mémorisé. Cette application et particulièrement les modèles biologiquement réalistes qui en découlent, présentent alors un modèle adaptatif de représentation distribuée compatible avec la biologie et qui en réplique certaines conditions, comme la parcimonie du nombre simultané de cellules activées et son lien avec le nombre potentiel de motifs mémorisables [Meunier et Nadal, 1995; Brunel et Sergi, 1999]. Ce travail inspira donc que des comportements complexes pouvaient émerger de réseaux simples et que cette étude gagnait de la confrontation avec des modèles de physique statistique, introduisant par la même une renaissance du mouvement connexionniste dans les années 1980.
Malgré les efforts croissants d'une formalisation solide et l'émergence d'applications efficaces étendant ou alternatives aux solutions classiques (on pourra voir Haykin [1999]), les approches connexionnistes ont souffert d'une manque de reconnaissance par rapport aux approches purement logiciennes car le résultat de leurs applications peut, dans leur ressemblance avec la "souples-

se" de la biologie, sembler trop imprévisibles et leurs codages et apprentissages ne se pliaient pas aux règles standardisées reflétées dans l'organisation de l'ordinateur de von Neumann. Paradoxalement, c'est aussi sûrement la nature de la modélisation initiale du neurone formel de McCullough et Pitts [1943] comme un système linéaire et suivant des mécanismes similaire aux calculateurs de l'époque (synchronie, poids statiques), que les réseaux de neurones formels n'ont pu encore se rapprocher de l'efficacité des systèmes biologiques.

Une nouvelle ère ?

Toutefois, une nouvelle ère semble se dessiner avec l'intégration du connexionnisme, avec l'Intelligence Artificielle, dans les Sciences Cognitives. L'interdisciplinarité et les échanges soutenus avec les neurosciences sont des facteurs importants d'unification d'une théorie de la vision et d'assouplissement des dogmes et frontières de chaque discipline. Il permettent aussi d'envisager des problèmes, comme celui qui nous intéresse, non plus seulement à un seul niveau de description (le neurone, la segmentation des objets, la sensation visuelle), mais de confronter ces approches sur plusieurs niveaux à l'image de la richesse des processus dans le SNC.
Par conséquent, l'étude des réseaux de neurones, les *neurosciences computationnelles*, est de fait devenue une discipline novatrice réunissant mathématiciens, informaticiens, physiciens et biologistes, permettant de confronter et comparer par des approches parallèles les méthodes classiques avec des modèles alternatifs. En étudiant la biophysique des calculs, ce mouvement participe ainsi, en collaboration avec d'autres champs des Sciences Cognitives, à la définition de modèles du fonctionnement des processus cognitifs.
Les récentes avancées de ce champ se concentrent en particulier sur l'aspect temporel des calculs neuronaux. Prévoyant ainsi l'émergence de nouvelles technologies de calcul, comme les calculateurs asynchrones, les systèmes distribués ou encore les systèmes d'exploitation (OS) non-monolithiques, ces nouveaux codes vont déranger beaucoup d'a priori. À l'image d'Internet, considéré par certains comme un "calculateur distribué planétaire", l'impor-

tance des calculs distribués et asynchrones vont ouvrir de nouveaux besoins et de nouvelles voies de recherche vers la modélisation de processus cognitifs qui accéléreront sans nul doute le rapprochement entre technologie et biologie.

1.1.3 L'inspiration biologique en traitement de l'image : vers l'ingénierie neuromorphique

En parallèle avec ces avancées théoriques, les applications et la technologie jouent un rôle primordial dans l'orientation des recherches dans les modèles de vision.

Applications de l'étude de la Vision

En effet, les applications utilisant les dernières avancées issues des neurosciences computationnelles ont une part grandissante dans les marchés des technologies de l'information. On peut citer le cas de la société Visionics, menée par J. Atick, dont les systèmes de reconnaissance de visages ont été choisis en particulier dans de nombreux aéroports américains. Cette société a fusionné récemment avec Identix pour devenir le leader mondial de la biométrie.
Aussi, ce champ touche alors tous les systèmes de traitement d'images d'un type plus général que celles auxquelles nous sommes habitués, comme les images satellitaires, et qui peuvent donc être composées de nombreuses bandes de fréquence ou encore transmettre des informations supplémentaires de phase (comme pour la technologie SAR). Des techniques neurales utilisant des Fonctions à Base Radiale ont par exemple été développées à l'ONERA, centre de Toulouse par Bruno Pelletier et permettent d'identifier les types de plancton à partir d'une telle "image" multibande.
Ces applications peuvent aussi recouvrir d'autres champs de recherche et le nouveau standard émergeant de compression d'image Jpeg2000, basé sur la théorie des ondelettes, est alors un exemple de fusion de techniques de traitement logique avec une structure mathématique inspirée de la biologie. De la même façon, les logiciels de la firme Synaptics qui "voient" les mouve-

ments du doigt sur le “trackpad” des ordinateurs portables* sont basés sur des techniques neuronales.

Ingénierie Neuromorphique

Cette dernière application est basée sur les principes naissants de l'*Ingénierie Neuromorphique* (IN). Les méthodes qu'elle introduit vont avoir une grande influence sur notre vision des processus cognitifs. En effet, l'IN a été définie par Mead [1989] comme un champ de l'ingénierie copiant l'architecture et la structure de systèmes de calculs sur les systèmes nerveux biologiques et est donc le complémentaire naturel pour les implantations physiques des principes des réseaux de neurones (voir Fig. 1.1.2). En particulier, son but vise à modéliser des processus, comme la transformation rétinienne, par des microprocesseurs dédiés qui approchent l'architecture et les performances des systèmes biologiques en insistant sur la physique des calculs et du code, tout en veillant aux contraintes comme la consommation électrique d'un système. Cette méthode revient donc à une *ingénierie inverse* qui, à l'image d'un développeur décompilant et désassemblant un programme pour en copier les rouages, tenterait de retrouver la représentation et le fonctionnement de l'information neurale. Le rôle de l'ingénieur neuromorphique est donc de “casser” ce code.

Ainsi, des applications comme l'implantation de réseaux de neurones dans des circuits intégrés de type aVLSI permettent de développer des applications en temps réel des modèles que nous allons présenter et introduisent des contraintes et des standards parallèles à ce qui s'observe dans la vision biologique. Ainsi, si l'on veut construire une caméra qui consomme le moins d'énergie possible tout en gardant de bonnes performances, on voit émerger dans les laboratoires d'ingénierie neuromorphique les premières rétines de silicone fonctionnelles avec des protocoles différents des normes industrielles [Boahen, 1997]. Ces applications nécessitent alors l'introduction de nouveaux cadres théoriques et on voit donc apparaître de nouveaux standards de communication point-à-point pour des puces électroniques comme le standard

*40% du marché

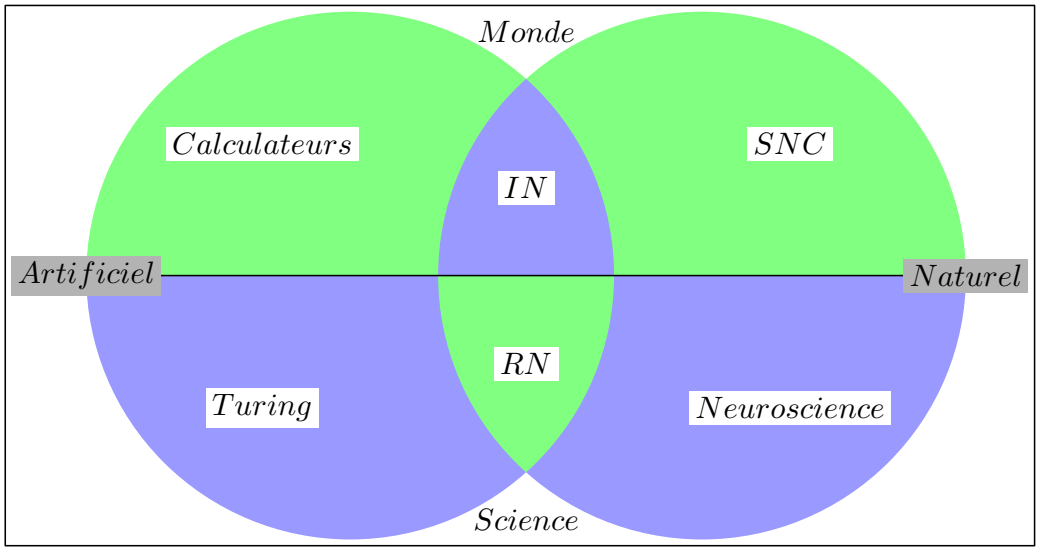

FIG. 1.1.2 – *Approches sur la nature des calculs dans les processus cognitifs.* Cette figure synthétise différentes approches physiques du calcul lié à la cognition. Elle fait apparaître leur dimension artificielle (à gauche) ou naturelle (à droite) et sépare alors les processus relevant du monde réel (en haut) (des calculateurs, comme ceux à transistors, au Système Nerveux Central (SNC)) de ceux relevant du monde de la connaissance (en bas) (des calculateurs théoriques, les Machines de Turing à l'étude scientifique du SNC, les neurosciences). Ainsi, les réseaux de neurones (**RN**) apparaissent alors naturellement à la croisée des neurosciences et des calculateurs théoriques comme l'étude théorique de calculateurs inspirés de la biologie. Enfin, l'Ingénierie Neuromorphique (IN) apparaît de façon complémentaire aux RN comme l'intermédiaire entre l'implantation de ces calculateurs et le SNC, développant ainsi une approche d'ingénieur sur les processus cognitifs en se concentrant sur la physique des calculs biologiques (figure inspirée d'une conférence de Rodney Douglas).

Address-Event Representation (ou *AER*), basé sur des événements [Boahen, 2000].

Applications neuromorphiques et robots

Une application immédiate de l'ingénierie neuromorphique est l'implantation de ces systèmes sur des robots. Un exemple marquant est par exemple le cas des robots autonomes utilisant la vision comme moyen de guidage sur un sol inconnu, à l'image du *rover* Sojourner sur la surface de la planète Mars. Dans ce cas particulier, la "vision" était tridimensionnelle (par un balayage de l'environnement par un rayon laser télémétrique) pour permettre une relative autonomie entre les délais imposés par le temps de transmission des commandes de la Terre à Mars (environ 10 min).
Elles s'étendent aussi à des missions de téléchirurgie, qui font notamment apparaître un besoin d'interactivité entre la vision de l'opération réelle et les gestes téléguidés du chirurgien. L'étude de telles applications peut donc changer les modalités, buts et contraintes sur le système que nous étudions. Cette approche développe la créativité de tels modèles et la définition des tâches qu'ils doivent accomplir et montre que notre vision de la vision doit être modifiée.
Nous voyons alors que la vision n'est pas un sujet d'étude évident, tant par sa généralité que par les processus qui la sous-tendent : la découverte du code de la Vision reste un défi.

1.2 Vers une vision dynamique

Après cette courte synthèse de l'état de l'art sur l'étude de la vision et de ses applications, nous allons souligner les aspects particuliers de la vision que nous allons explorer. En particulier, notre but va être de définir les aspects de la vision qui permettent de mieux la comprendre comme un processus dynamique et adaptatif. Nous allons donc confronter des faits sur la structure, la fonction et le caractère adaptatif des processus visuels pour soulever une problématique sur le code de la Vision.

1.2.1 Quel est le rôle du temps dans le calcul neural ?

Tout d'abord, en étudiant les principes qui nous permettent d'envisager les mécanismes neurophysiologiques comme l'implantation biologique d'un calcul neuronal, nous allons étudier la place que prend la variable temps dans le code visuel.

Existe-t-il un "langage" dans l'activité neurale ?

Il est d'autant plus surprenant pour le novice que pour un neuroscientifique d'observer des enregistrements de l'activité cérébrale, ou encore de pouvoir entendre le son produit dans un haut-parleur en amplifiant cette activité. On observe avant tout, sur l'Électro-EncéphaloGramme (EEG) par exemple, que plongée dans un bruit de mitraille, cette activité suit des rythmes et fluctuations complexes. Toutefois, malgré une grande disparité individuelle, l'activité globale moyenne des neurones reste relativement homogène autour de 1 à 10 Hz. On peut alors suivre autour de cette moyenne des "vagues" mais qu'on ne peut pas relier explicitement à l'activité réelle d'une fonction comme pour l'Électro-CardioGramme (ECG), par exemple.
En fait, ce qui semble être l'artefact d'un "langage neural" est interprété comme les variations d'activité des neurones mais aussi comme les oscillations des champs électriques locaux que ces activations provoquent. Ce "langage" suit des chemins spécifiques et est particulièrement sensible à la succession exacte des événements. Il a par exemple été mis en évidence que dans une tâche de catégorisation entraînant l'activation d'une aire spécifique du SNC, si l'on dérangeait la temporisation de l'activité dans cette aire (sans changer sa réponse mais en la retardant suivant un délai précis), on pouvait bloquer la réalisation globale (ici la prise de conscience) de cette tâche. La vision est donc un processus dépendant au niveau de son organisation macroscopique de l'ordre de la temporisation relative des activités de ses sous-modules : la vision est donc au niveau macroscopique un processus *asynchrone*.
Aussi, les expériences de Thorpe et al. [1996] ont montré qu'il existait des contraintes sur le temps de traitement visuel. Ces expériences, que nous décrirons en détail dans la Sec. 2.3, montrent que la vue traditionnelle sur le

code neural est incompatible avec les observations biologiques et exige dans ce cadre une nouvelle exploration de la physique de ces calculs. Par exemple, une part de l'information doit pouvoir être fournie exclusivement en avant pour privilégier le temps de réponse : le chemin le plus court entre l'œil et le cortex moteur pour ce traitement rapide est donc direct (il ne "boucle pas). Ce modèle remet alors en cause de nombreux modèles établis et permet d'émettre des hypothèses nouvelles sur les modèles candidats qui puissent implanter un tel traitement.

La synchronie et l'asynchronie

On interprète souvent le code au niveau des sous-parties du SNC comme des *synchronisations* temporelles à différents niveaux (entre neurones, groupes ou aires). Pour répondre aux expériences de catégorisation ultra-rapide, nous allons resserrer notre niveau d'étude à des *assemblées de neurones* que nous essaierons d'étudier de façon générique. À ce niveau, la synchronisation est compatible avec le problème de *liage* de différentes modalités et permet alors d'imaginer que le concept de "pomme rouge" puisse être représentée temporellement par la synchronisation de deux assemblées de neurones respectivement sensibles aux pommes et à la couleur rouge. Toutefois, on peut douter tout d'abord de l'existence de l'implantation biologique du liage comme une synchronisation, mais aussi se demander si la représentation correspondant à cette approche largement acceptée est satisfaisante. En effet, celle-ci n'est ni optimale, ni compatible avec un traitement rapide et on peut être tenté de trouver des alternatives à ce code.
En effet, si la synchronie semble jouer un rôle important, elle peut n'être qu'un aspect d'un code plus complexe. Ainsi, par exemple, la précision du code lié à la localisation chez la chouette par des informations auditives stéréoscopiques fait intervenir des délais temporels relatifs inférieurs à la milliseconde [Gerstner et al., 1996]. Ceux-ci sont sûrement interprétés par un code plus riche qu'une simple détection de coïncidence. De plus, si l'on s'accorde à dire que les mécanismes de compétition sont essentiels pour la richesse du codage neural, on peut se demander si ces mécanismes, en dehors

du mécanisme où les gagnants d'une telle "élection" prennent tout, ne font pas plutôt intervenir des résultats modulés suivant le résultat. Comment ces réponses sont alors codées puis interprétées par un code impulsionnel ? Quel est alors le lien avec le fonctionnement asynchrone des assemblées de neurones ?
Ces questions sont primordiales car elles touchent à la vraie nature du code neural (et du code visuel en particulier) comme nous l'avons défini plus haut. Si classiquement, on imagine un neurone comme un opérateur logique ou un calculateur primitif, on peut —en étendant l'idée originale de McCullough et Pitts [1943] dans le contexte des calculateurs à relais— l'imaginer comme un système complexe modulant et aiguillant dans le temps les signaux impulsionnels à travers les chemins qui constituent les réseaux du SNC. Dans ce cas, le neurone devient un nœud adaptatif dans une assemblée temporaire que l'on ne peut dissocier de l'ensemble sans lui faire perdre son sens. Nous proposons ainsi que l'aiguillage de l'information, contrôlé par une connectivité adaptative précise, est alors à l'origine des processus cognitifs.

Les neurones à impulsion

Dans ce cadre, les impulsions neuronales semblent clairement constituer le support universel de l'information dans le SNC et l'étude des mécanismes neuronaux vont nous aider à mieux comprendre comment baser le code neuronal sur ces impulsions. En effet, ces pics d'intensité électrique (de moins d'une ms pour environ 0.1 V) déclenchent les événements de transmission neuronale (au niveau des synapses) ou d'action (au niveau des muscles par exemple), d'où leur dénomination de *potentiels d'action (PAs)**. De plus, ils sont quasi-généralisés dans les SNC des êtres vivants et sont pour un individu très semblables. Ces aspects se rapportent alors de nouveau à un niveau plus détaillé du code neuronal et qui conduit à étudier le neurone comme un système complexe (métabolisme, communications) que l'on peut encore décomposer en sous-systèmes, et dans lequel les synapses semblent constituer un rôle clé.

*ou *spike* en anglais

En particulier, nous allons étudier quel est le corrélât de l'information qu'ils véhiculent à travers ces impulsions. En effet, ces impulsions montrent que le code neural est discrétisé en amplitude par l'arrangement spatio-temporel des impulsions. Cet aspect discret qui se rencontre rarement dans la nature, est analogue au codage de l'information génétique par l'ADN et ses expressions. En effet, le support de l'information génétique permet de "traduire" la synthèse de tout un "dictionnaire", chimique et spatial, de molécules avec un nombre limité d'acides aminés, qui s'assemblent préférentiellement suivant les "mots" (ou *génons*) créés par la combinaison des bases. Certaines hypothèses imaginent alors qu'il existe de même des "mots neuronaux" de base et des lettres pour l'alphabet du langage neural qui s'assemblerait suivant une grammaire spécifique. Toutefois, cette "langue", si elle existe, n'a pas été décryptée, et on se demande plus simplement quelles peuvent être des règles de composition de ces signaux impulsionnels et comment expliquer par exemple la grande irrégularité du signal impulsionnel des neurones [Softky et Koch, 1993a,b].
Pour bâtir un modèle de la Vision, il faut donc définir un code en imaginant des règles plausibles de composition et de transmission de l'information basées sur ces impulsions. Les synapses sont alors des sous-systèmes neuronaux essentiels de ce code, loin de l'approche symbolique de leurs modèles dans les neurones formels. Ces idées nous pousseront à développer des algorithmes basés non sur un temps continu mais sur les *événements* neuronaux et synaptiques (à commencer desquels l'impulsion neuronale) qui constitueront les bases d'un code visuel impulsionnel.

1.2.2 Quels modèles de codage impulsionnel rapide ?

Les mécanismes cognitifs rapides que nous avons évoqué montrent que le SNC obéit à des processus temporels qui sont incompatibles avec les théories classiques du code neuronal. Peut-on alors proposer en utilisant l'information impulsionnelle des modèles de construction d'un code neuronal adapté à ce régime particulier du fonctionnement neuronal ?

Le calcul neural, vers un code événementiel

Concentrons-nous sur la nature du calcul neural. En effet, le SNC possède une capacité de calcul immense grâce à sa structure parallèle [Thorpe et Imbert, 1990] mais présente des limites quant à sa connectivité (les connexions entre neurones prennent à elles seules 95% du volume du cerveau) alors qu'un calculateur classique monoprocesseur, à l'inverse, peut communiquer très rapidement mais possède une capacité de calcul relativement faible. Comment imaginer alors un autre paradigme sur le calcul neural que celles inspirées des calculateurs actuels ?
En particulier, l'expérience neuropsychologique de Thorpe et al. [1996] que nous avons citée plus haut permet de proposer un code dans une assemblée par l'ordre d'activation des neurones. Cette proposition iconoclaste, qui semble a priori incompatible avec les principes biologiques, permet de construire une théorie neurale, basée sur cet axiome et qui conduise à une nouvelle "mathématique" du code neuronal qui exhibe de très bonnes performances par rapport à d'autres types de codage. De plus, des modèles biologiques permettent de la justifier mais aussi d'en préciser la définition.
La vision pourrait alors dans ce cadre être décrite par un flux d'événements "crypté" par les chemins qu'il emprunte, selon des "clés" comme celle de l'ordre d'activation des neurones sur une aire. Les neurones —considérés alors dans ce type de réseaux comme des "routeurs"— utilisent ce code pour savoir s'il doivent relayer cette information et avec une mesure modulée par leurs temps de tir relatifs.

Adaptation, écologie et évolution dans les neurones impulsionnels

Ce type de code basé sur le rang implique que l'information est codée grâce aux coopérations et compétitions entre neurones. La réponse des différents mécanismes neuronaux doit donc être adaptée pour que ces compétitions optimisent l'information portées par les PAs tout en respectant les contraintes métaboliques. Par exemple, si l'on simplifie le nerf optique à un canal bruité de transmission de l'information, des principes d'optimisation complémentaires viseraient à maximiser la qualité de la transmission de l'information tout

en minimisant la consommation d'énergie liée à cette transmission. Ainsi, par des principes de décorrélation de l'intensité de deux points voisins dans des images naturelles, Atick [1992] a pu proposer un modèle maximisant le transfert d'information dans le nerf optique et compatible avec les courbes de sensibilité observées dans la biologie [Enroth-Cugell et Robson, 1966].
Cette faculté d'adaptation prend une part importante dans la définition d'un code de la Vision. De façon plus large, on peut regrouper sous le terme d'*écologique* les facteurs qui permettent de prévoir de manière générale l'évolution de systèmes visuels par l'optimisation de leur efficacité. Celle-ci peut prendre en particulier la forme de principe de maximisation d'un but (par exemple la qualité de transmission de l'information pour Laughlin [1981]) ou similairement de minimisation d'une contrainte (la redondance du signal, par exemple [Barlow et Foldiak, 1989]). Ceux-ci permettent alors de déduire des comportements universels d'adaptation des neurones que l'on peut traduire dans le cadre d'un code visuel impulsionnel. Toutefois, les méthodes écologiques ne sont pas des approches téléologiques, mais permettent de déduire l'architecture du système en l'étudiant comme un produit de l'Évolution.
Aussi, il semble alors que ces principes écologiques soient des facettes de principes adaptatifs unitaires. Différentes approches d'optimisation de la représentation neuronale se rapprochent en fait de l'émergence de codes détectant de façon non-supervisée la *cohérence* du signal, c'est-à-dire capable de séparer de façon aveugle les sources du signal. Ces facultés d'adaptation coexistent alors à plusieurs échelles temporelles. On peut par exemple étendre ces mécanismes à l'évolution à plus long terme du système visuel et expliquer comment, alors que le système visuel humain est encore en développement à la naissance, la majorité des fonctions visuelles (comme la stéréopsie ou le suivi de mouvements) sont déjà stables chez un enfant de 6 mois.

Implantation du code visuel rapide dans le SNC

Dans notre recherche du code neural rapide et de son implantation physiologique, il est essentiel de circonscrire les fonctions de ce code. Quel est par

exemple le rôle de la rétine, cette seule partie du cerveau qui en relativement détachée, et dont la sortie, le nerf optique, constitue un véritable “goulot d'étranglement” pour l'information visuelle ? Son étude physiologique révèle que la transformation de l'information lumineuse en un train d'impulsions peut avoir plusieurs fonctions découplées dans le temps comme transmettre l'information au plus vite ou comprimer la taille d'une représentation fine de la scène visuelle.Le nombre limité de fibres dans le nerf optique peut alors être une contrainte majeure de fonctionnement.
En passant la rétine, d'autres fonctions semblent apparaître dans les aires visuelles primaires qui sont situées dans le cortex occipital. En particulier, l'aire visuelle primaire voit une explosion du nombre de neurones par rapport au nombre de fibres arrivant de l'œil, ce qui entraîne l'explosion du nombre de représentations et transformations possibles de l'information visuelle. On dit alors que la représentation est *sur-complète*. On pourra alors décrire ces étapes du traitement visuel par une architecture parallèle et hiérarchique.
Enfin, en s'approchant des aires motrices et du cortex préfrontal, ce flux doit pouvoir permettre de catégoriser l'activité pour fournir une réponse adaptée. On peut se demander s'il existe alors une stratégie optimale pour fournir une réponse rapide et quel est le meilleur compromis entre vitesse et précision. Le contexte joue alors un rôle primordial, et la réussite de la réponse conditionnent alors l'adaptation du système et la propagation du flux de l'information visuelle. On peut alors finalement relier le fonctionnement neural induit par la stimulation lumineuse à des buts comportementaux.

1.2.3 Vers des modèles fonctionnels de la vision

Ces modèles, que l'on implantera le plus souvent grâce à des simulations, ne doivent pas nous faire oublier les contraintes physiologiques liées à une implantation physique de ces calculs. Alors que le premier calculateur, l'Eniac, consommait 174 kW peut-être pourra-t-on grâce à des systèmes neuromorphiques alors rapprocher les ordinateurs actuels (consommant environ 100 W) de l'efficacité du cerveau humain qui ne consomme que l'équivalent de 10 W. Pour cela, il est nécessaire d'implanter des modèles fonctionnels du

code visuel tout en s'inspirant du fonctionnement du système visuel.

Contenu de l'information neuronale

En estimant physiologiquement le nombre minimum de neurones nécessaires pour coder un "percept", Kreiman et al. [2002] se sont aperçus que des neurones pouvaient répondre explicitement et exclusivement à des sensations complexes, une capacité surnommée "neurone grand-mère" dont l'exemple le plus marquant dans cette expérience est alors le neurone "Bill Clinton" découvert chez un patient. Ce neurone présentait une activité si et seulement on lui présentait une photographie ou un dessin de l'ancien président des États-Unis d'Amérique. Peut-on alors considérer que les représentations neuronales sont de ce type ? Sont-elles vraiment efficaces et quelle est l'étendue de la généralisation de cette représentation ?
On observe alors que, en conjonction avec cette représentation qui utilise un "dictionnaire" très large, les réponses des neurones révèlent en général un caractère *épars*, c'est-à-dire que dans une assemblée, peu de neurones sont activés en même temps. Est-ce alors une contrainte physiologique (*i.e.* d'économie) du calcul parallèle des neurones ou plutôt une stratégie efficace de codage et de représentation de l'information neurale ?
Aussi comme ces représentations sont très nombreuses, elles sont inter-dépendantes et leurs sensibilités se recouvrent. On observe alors —en particulier dans le cortex— que ces représentations sont fortement corrélées à leur implantation physiologique, mettant en avant l'importance des interactions latérales. L'étude d'implantations possibles de ces mécanismes nous éclairera alors sur les fonctions possibles de ce code impulsionnel épars.

Auto-organisation

Aussi, l'étude de l'anatomie du SNC, et du système visuel en particulier, révèle une grande précision de son architecture. Comme il est exclu que ces principes soient seulement supervisés (par le code génétique par exemple), quels sont alors les principes qui les guident et peut-on alors les relier à des buts et fonctionnalités ? Peut-on suggérer l'existence de processus d'auto-

organisation ?
On peut alors imaginer que les représentations du type de “Bill Clinton” vont émerger de façon non-supervisée dans des assemblées de neurones. Toutefois, on peut se demander quels sont les principes sous-jacents guidant la formation de telles assemblées, c'est-à-dire savoir quels principes adaptatifs reliés avec ce mode de représentation épars font émerger certaines catégorisations plutôt que d'autres. En particulier, peut-on alors relier ces mécanismes d'apprentissage non-supervisé et d'auto-organisation à des processus neurophysiologiques et une modélisation qui les implantent ?

Principes unitaires d'un code événementiel

On a vu que l'impulsion neuronale constituait un “quanta” d'information dans le SNC. Existe-t-il alors d'autres principes unitaires qui guident le flux d'information depuis l'œil jusqu'à une réponse motrice ? L'impulsion neuronale constitue-t-elle le seul *événement* de l'information ?
Ainsi, l'anatomie du cortex observe des similarités structurelles, qui suggèrent que les différences fonctionnelles proviennent de principes unitaires. En particulier, les *colonnes corticales* semblent constituer un système autonome dont on peut extraire une formalisation (en terme de communications neurales et de l'état interne de la colonne). Ce système peut alors être répliqué sur tout le cortex comme une brique élémentaire de construction. Un tel modèle —dont on réduirait le nombre de paramètres qui règlent son comportement et l'architecture de ses interactions (locales ou distales)— pourrait alors permettre de construire un code complet du traitement visuel, basé sur des événements produits par les colonnes corticales. Ces questions doivent alors nous guider vers une théorie unifiée reliant l'aspect temporel et impulsionnel du code dans une objectif d'une implantation d'une *vision dynamique*.

Le but de la thèse est donc bien défini. En nous fondant sur l'aspect temporel des processus neuronaux et sur la neurophysiologie du système visuel, nous allons étudier des modèles de codage impulsionnel. Ceux-ci vont nous permettre de définir des algorithmes de *codage dynamique* compatibles avec la rapidité du traitement visuel. De plus, nous allons aussi explorer

des représentations efficaces de l'information visuelle et intégrer des règles d'adaptation en utilisant des principes écologiques basés sur la statistique des images naturelles. Nous explorerons alors l'implantation d'un modèle de représentation sur-complète qui va ouvrir la voie à l'implantation d'un codage impulsionnel épars et de systèmes visuels adaptatifs complets pour proposer ainsi un modèle dynamique et écologique du code visuel.

1.3 Organisation de la thèse

Une fois le sujet défini, il est alors nécessaire de préciser les moyens utilisés en pratique durant la thèse mais aussi le style adopté et un plan détaillé du manuscrit.

1.3.1 Méthodologie scientifique

Catégorisation Ultra-Rapide

Les idées que nous allons exposer dans ce travail ont largement été inspirées par les travaux deThorpe et Imbert [1989] et Thorpe et al. [1996]. Elles prétendent donc ouvertement soutenir ces travaux pour étudier des alternatives computationnelles compatibles avec les résultats de ces expériences et en étendant les modèles et codes proposés par une étude mathématique approfondie. Nous porterons donc en particulier notre attention sur le codage temporel ultra-rapide de l'information visuelle.
Dans ce cadre, nous restreindrons le domaine de notre étude aux conditions de ces expériences que nous expliciterons à la Sec. 2.3, c'est à dire que nous étudierons la vision pour des images statiques "flashées". De plus, ces images seront toujours en niveaux de gris et nous nous intéresserons particulièrement au comportement rapide de la vision et donc à l'information portée en avant par la volée d'impulsions produite par l'image. Pour simplifier l'influence du contexte, nous considérerons aussi dans les modèles que l'image est présentée après un temps d'adaptation qui permet de considérer le système visuel au "repos".
Nous nous attacherons de plus à montrer que les modèles alternatifs proposés,

comme le codage de l'information impulsionnelle par l'ordre de décharge, sont des processus généraux, une "philosophie" que l'on peut étendre avec le cadre relativement abstrait de leur formalisation mathématique actuelle.

Simulations numériques et applications robotiques

Les images utilisées proviennent de prises de vues personnelles en extérieur ou en laboratoire. Alternativement, on a utilisé des bases de données publiques pour permettre une comparaison avec des résultats existants. En particulier, les comparaisons avec d'autres algorithmes sont effectuées si possible sur les mêmes images.
Les modèles que nous allons présenter seront appuyés par des simulations numériques. Celles-ci ont été écrites en C (*LCC* pour WINDOWS© et *GCC* sous Unix) et grâce au logiciel commercial de calcul et visualisation MATLAB©, sous différentes plates-formes*. Ces modèles profitent aussi des algorithmes écrits pour SPIKENET©. De façon générale, les simulations sont présentées sur les figures et détaillées dans le texte. Les scripts commentés correspondant à ces modèles (en particulier à chaque figure qui correspond à une simulation) sont accessibles afin de garantir la reproductibilité des figures présentes dans la thèse.
Les applications robotiques ont principalement été simulées dans un premier temps à partir de d'images enregistrées depuis des situations réelles et sont disponibles sur Internet. Complémentairement, des simulations de perception active ont été menées au Laboratoire Informatique et Automatique de SUPAÉRO, Toulouse et en collaboration avec D. Martinez au LORIA.

Style d'écriture et conseils de lecture

Ce manuscrit se veut largement interdisciplinaire. Pour garantir la lisibilité, nous avons privilégié le recours massif à des définitions et de notations unifiées pour éviter les confusions. Une table d'abréviations et les hyperliens de la version électronique vise à respecter la fluidité du texte. Nous allons

*Le logiciel sous licence GPL *Octave*, permet dans une large mesure de réaliser ces simulations. De même le logiciel *SciLab* permet lui de convertir les scripts dans un langage similaire.

aussi essayer d'exploiter la contrainte de langue de manière constructive pour proposer de normaliser en français des termes consacrés en anglais. Nous indiquerons alors les néologismes quand ils sont inhabituels.

1.3.2 Plan de l'étude

Comme il a été suggéré, ce travail va débuter par l'étude de la neurophysiologie du système visuel pour ensuite proposer des modèles de neurones et de réseaux neuronaux de codage rapide. Nous proposerons alors un modèle de codage impulsionnel épars et nous présenterons aussi des applications de ce modèle.

La neurophysiologie du système visuel

Dans un premier temps, le chapitre 2 présentera les fondements physiologiques des processus de la vision. Tout en gardant une vision d'ingénieur, nous soulignerons en particulier les aspects temporels de la biophysique des calculs dans le système visuel et nous étudierons l'organisation distribuée et les particularités du traitement de l'information lumineuse.
En effet, nous nous concentrerons dans une première section (voir Sec. 2.1) sur les aspects dynamiques du calcul neural, en redéfinissant précisément le but de la vision, la physique des calculs biologiques puis enfin en décrivant le flux d'information dans le SNC. Cette section s'étendra par une étude de l'anatomie du système visuel, depuis la rétine (voir Sec. 2.2.1) jusqu'aux aires visuelles supérieures, pour enfin étudier les modèles neuromimétiques de représentation de l'information visuelle (voir Sec. 2.2.3).
Cette étude pourra alors introduire en détail les expériences de Catégorisation Ultra-Rapide (CUR, voir Sec. 2.3) et leurs conséquences sur nos modèles de la vision. Cette approche neurophysiologique pourra alors s'achever sur une définition précise des contraintes imposées par la CUR pour proposer des modèles alternatifs de la vision.

Modèles de codage rapide

Nous étudierons alors dans le chapitre 3 des modèles computationnels de ce codage rapide. Tout d'abord, nous analyserons des modèles neuronaux de codage impulsionnel (voir Sec. 3.1). En étudiant des mécanismes détaillés et leurs conséquences sur le codage, nous proposerons des mécanismes d'apprentissage pour le neurone mais aussi pour une assemblée de neurones. Ces résultats nous aideront alors à établir les bases d'un code impulsionnel par la latence (voir Sec. 3.1.3).
Nous pourrons alors analyser le modèle de rétine présenté par Van Rullen et Thorpe [2001]. Celui-ci aApropose un codage impulsionnel de l'information rétinienne, et en particulier un codage de cette information par l'ordre d'activation (voir Sec. 3.2.2). Nous étudierons mathématiquement cette transformation et proposerons des alternatives constructives aux problèmes rencontrés. En particulier, nous proposerons une optimisation statistique de la propagation de ces impulsions grâce aux statistiques des images naturelles (voir Sec. 3.2.3) que nous relierons à une distance subjective entre les images. Cette dernière section nous amènera alors naturellement à une formalisation du codage impulsionnel par l'ordre d'activation (voir Sec. 3.3). Après avoir présenté un modèle biologique plausible, qui nous éclairera sur ces principes, nous étudierons cette transformation de façon formelle en utilisant les outils des statistiques non-paramétriques (voir Sec. 3.3.2). Des applications illustreront ces résultats et montreront en particulier l'importance des paramètres de cette formalisation mais aussi d'une définition exacte du rang.

Vers un code dynamique épars

Après avoir évoqué des limites du système précédent, nous étudierons des modèles d'interactions latérales qui étendent les possibilités de ce code neural. En effet nous pourrons alors utiliser une *représentation sur-complète* de l'information (voir Sec. 4.1.1). Dans un premier temps, nous étudierons un modèle de rétine avec une telle représentation en la comparant avec le modèle précédent. Ce modèle sera basé sur un algorithme de Poursuite de Concordance (Matching Pursuit en anglais), et nous verrons comment cet

algorithme s'adapte particulièrement bien au codage impulsionnel rapide.
Cette étude se prolongera par la définition d'un modèle de l'aire visuelle primaire par un code impulsionnel basé sur une représentation sur-complète de filtres sélectifs à des orientations. Nous présenterons succinctement son architecture comme un modèle complet de colonne corticale et nous étudierons les applications découlant de cette représentation en tenant compte d'une connaissance a priori. De plus, nous définirons des règles non-supervisées d'apprentissage que nous comparerons avec le modèle de Olshausen et Field [1996]. Nous présenterons alors les résultats de ce codage en terme de représentation de l'information et de sa rapidité, et discuterons de son optimisation statistique. Nous l'étendrons en particulier à l'étude d'un système d'extraction de primitives dans l'image (comme les contours). Aussi, nous comparerons ce système au standard de compression JPEG et montrerons que notre modèle permet la transmission rapide et progressive de l'information visuelle avec une qualité comparable pour de forts taux de compression.
Enfin, nous présenterons des applications du codage impulsionnel épars que nous combinerons avec le modèle de rétine que nous avons présenté et en particulier, nous montrerons comment ce principe peut être implanté pour être étendu à d'autres couches. Nous étudierons alors l'efficacité de ce modèle hybride et déduirons un modèle simple d'attention dans ce modèle de propagation que nous comparerons à l'attention visuelle biologique et au modèle de Itti et al. [1998]. Enfin, nous évoquerons des applications robotiques complètes en étudiant tout d'abord un système neuromorphique combinant un modèle d'attention avec une autre modalité sensorielle puis en décrivant une application pour un robot mobile à guidage visuel dont la tâche est de détecter une cible connue.

Chapitre 2

Neurophysiologie du système visuel

> Learning the secret of flight from a bird was a good deal like learning the secret of magic from a magician. After you know the trick and what to look for, you see things you didn't notice when you did not know exactly what to look for.
>
> *Orville Wright*, 1913

2.1 La vision comme un système dynamique

La vision est une sensation si familière qu'il est parfois difficile d'imaginer la complexité des processus soutenant la simplicité de sa perception. Pour la définir comme un processus dynamique, nous allons devoir déranger le sens commun et redéfinir des principes fondamentaux, comme son but, son corrélât neuronal et son fonctionnement.

2.1.1 Quel est le but de la vision ?

Nous allons essayer de définir précisément les concepts que nous avons introduits en Sec. 1.1 pour définir le but de la vision.

La vision est une stratégie

La vision est une stratégie chez les êtres vivants pour interagir avec le monde lumineux. Ce monde lumineux est constitué des photons qui forment les rayons lumineux et émis ou réfléchis par les objets du monde et qui constitueront donc une source d'information sur le milieu. L'organe sensible de la Vision, l'œil, permet en focalisant les rayons optiques de concentrer une partie de cette information sur sa partie photo-sensible, la *rétine*, en formant une image* du monde lumineux. Une partie de cette information (en particulier seules les ondes entre 400 et 700 nm, la lumière visible) va être codée pour être traitée dans le SNC. On peut alors découpler un sous-système du SNC spécialisé dans ces tâches, le système visuel† qui permet alors d'analyser spécifiquement cette information. Ainsi, la vision va participer en parallèle avec les autres sens et fonctions cognitives aux processus vitaux d'une façon progressivement de plus en plus intégrée.
Le but de la Vision, dans le sens où nous l'avions défini à la Sec. 1.1.1, est de fournir une réponse comportementale pour agir, planifier ou surveiller une action qui s'intègre dans les processus vitaux. Or la diversité et l'intégration des modalités perceptives suggèrent alors que les différents processus cognitifs partagent un "langage" commun et donc que le SNC est un système modulaire dont les sous-parties partagent des mécanismes similaires. La vision doit donc créer —entre la sensation et la réaction— une *représentation* nécessaire à l'intermodularité. Cette représentation doit pouvoir être communiquée avec les systèmes moteurs (agir), limbiques (interactions hormonales) ou internes (associer et apprendre) tout en intégrant les contraintes extérieures ou physiologiques (comme la consommation énergétique du cerveau). Le but de la vision peut donc être défini comme l'ensemble des processus qui de l'image du monde conduit à une représentation qui puisse être ensuite utilisée par les différents modules d'action.
Mais alors, de nombreuses contraintes modifieront le but de la vision. À

*On définit ainsi de façon générale une image comme un ensemble d'informations lumineuses organisée spatialement sur une surface.

†Ses frontières dépendent de sa définition exacte, mais est toutefois clairement définie jusqu'aux aires visuelles secondaires.

titre d'illustration, concentrerons-nous ici sur le cas d'un humain à la recherche d'un visage connu. Le but sera alors différent si la contrainte consiste à donner la réponse la plus rapide ou la plus précise. De plus, les conditions (jour ou nuit, direction de l'éclairage) et le contexte (danger immédiat, non-attente de la stimulation) sont encore autant de facteurs qui changent alors la représentation idéale de l'information et donc le but des processus visuels. La situation peut encore être différente si l'on peut se mouvoir et donc se diriger vers le visage recherché : la vision devient alors une stratégie exploratoire active. L'ensemble de ces buts vont alors influencer l'évolution de la structure du système visuel.

Implantations de la vision dans l'Évolution

La diversité de ces buts se reflète dans les différentes formes de la vision émergeant de l'Évolution, depuis le mécanisme neuromoteur phototactique de l'algue unicellulaire *euglena gracilis*, via les yeux composés de nombreux insectes comme la mouche, jusqu'à la vision active et consciente des humains et d'autres mammifères. Cette diversité montre aussi que cet inventaire n'est pas en ordre croissant de qualité, puisque par exemple, l'on ne peut comparer la précision de la vision du faucon aux capacités tétra-chromates du pigeon. Une conséquence marquante est par exemple la position des yeux chez de nombreux mammifères, privilégiant la vue périphérique pour les proies (comme le lapin) et la vue en avant binoculaire et focalisée des prédateurs (comme l'homme). Cette diversité ne doit cependant pas cacher des principes similaires (comme l'unité des systèmes visuels des mammifères) dûs en grande partie au caractère continu et progressif de leur évolution.
En sautant aux primates, nous pouvons donc statuer que l'évolution va privilégier les modifications physiologiques qui amélioreront l'efficacité de la vision par rapport aux buts, à l'environnement ambiant et aux contraintes. Nous pouvons alors affiner notre définition à l'ensemble des processus permettant de générer un *code*, c'est-à-dire un ensemble de calculs transformant l'information sensorielle, qui puisse être intégré aux autres processus cognitifs. Ce code va ainsi permettre de transformer une représentation par vues en une

description structurelle (en informatique, d'une image *bitmap* à une image vectorielle) exploitable par les processus cognitifs supérieurs. En termes informatiques, la vision est un algorithme, une suite de calculs, amélioré en fonction des buts du système vivant par les processus de l'Évolution et utilisant le substrat organique du SNC.
Par extension, nous pouvons en conclure qu'un but de la vision est de créer un *modèle génératif* qui permette de créer ou reconnaître des représentations visuelles depuis le vécu de l'animal. Ce modèle génératif correspond à une stratégie pour s'adapter par exemple aux situations visuelles nouvelles. Le code de la vision doit pouvoir alors induire des représentations nouvelles depuis celles qui ont été apprises. Par exemple, les objets sont souvent en translation dans le monde physique et vont fournir des informations cohérentes dont le code doit pouvoir extraire les régularités au niveau des mécanismes et de la représentation correspondante. Ce code ne va pas apprendre la règle de translation comme une règle écrite mais va apprendre par des processus adaptatif un modèle génératif qui va par exemple pouvoir "interpoler" une information depuis des sources incomplètes.

La vision et les processus vitaux

Les processus participant à la vision doivent donc pouvoir s'adapter à l'environnement pour répondre au mieux à l'ensemble des processus vitaux. À long terme, le code génétique et la plasticité des processus de calcul dans le SNC sont deux exemples de stratégies du vivant pour évoluer et s'adapter à différentes échelles de temps. Ainsi le système visuel de l'embryon illustre parfaitement comment l'adaptation neuronale organise et intègre le système visuel en fonction du développement neural "programmé" génétiquement et de l'environnement. Par conséquent, les processus doivent pouvoir être évolutifs, c'est-à-dire qu'une solution qui ne permet pas d'adaptation aura plus de chance de disparaître même si elle est plus performante.
À plus court terme, il paraît alors évident qu'un rôle majeur de la vision est de garantir une perception stable et robuste en fonction des contraintes extérieures. Ainsi, s'il parait peu étonnant que nous puissions nous adapter

à des changements de contrastes, ce niveau d'adaptation (de 1 lumen[‡] pour une feuille blanche sous la lumière de la pleine lune jusqu'à 10 000 lumen pour une journée ensoleillée) est inégalé par la technologie actuelle. À plus long terme, on observe aussi que notre représentation se révèle constante malgré des changements importants. Dans un cas extrême, une perception inversée par des lunettes spéciales (par exemple où par des prismes le haut se retrouve en bas) s'adapte après quelques jours pour retrouver une perception "normale", c'est-à-dire cohérente par rapport à l'espace physique et qui permette d'effectuer des actions quotidiennes. Cette observation suggère que des principes homéostatiques sont primordiaux dans le système visuel et dans le SNC en général et qu'une exigence des processus visuels est donc de pouvoir être régulés pour s'adapter aux différents environnements rencontrés.
Finalement, on peut dans le cadre de notre travail définir la vision comme une stratégie, s'intégrant dans les processus cognitifs et donnant une représentation stable du monde. Son code, c'est-à-dire la représentation et l'ensemble des calculs visuels, doit correspondr à une stratégies optimale dans le compromis entre la construction d'une relation cohérente du monde physique et par rapport à l'ensemble des contraintes.

2.1.2 La vision et son corrélât neuronal

Pour mieux comprendre comment l'Évolution a implanté cet ensemble de processus, nous allons étudier de façon générale l'architecture des processus neuronaux. En particulier, nous insisterons sur les différents niveaux d'études du SNC depuis les cellules neuronales jusqu'au réseau complexe que ces cellules forment.

Le système visuel est un réseau de neurones

Depuis les travaux de Cajal [1911], nous savons que le SNC est composé d'un complexe réseau discontinu de cellules spécialisées, les *neurones*, soutenues par des cellules de soutien (principalement des cellules *gliales*, dans une proportion de 90% du nombre total de cellules). Le rôle de ces dernières

[‡]Le *lumen* est une unité mesurant la puissance lumineuse

cellules est de subvenir aux besoins des neurones : énergie, évacuation des débris, mais aussi de structurer le réseau des neurones. La membrane des neurones comporte des canaux ioniques spécialisés s'ouvrant ou se fermant pour laisser le passage (activement ou passivement) aux ions en solution au gré de l'activité locale. Ces canaux produisent une différence de potentiel du milieu extérieur avec l'intérieur hyperpolarisé et peuvent propager de proche en proche ce déséquilibre électrochimique*, qui peut ainsi constituer un signal neuronal.

Un neurone peut prendre des morphologies très différentes mais est fonctionnellement constitué de structures similaires : un prolongement richement arborisé, les *dendrites*, un corps cellulaire (ou *soma*) et un *axone*. On appellera alors *potentiel du neurone* la valeur du potentiel au soma. L'arbre dendritique reçoit les signaux des organes sensoriels ou d'autres neurones, les transforme pour enfin les transmettre au corps cellulaire. Celui-ci, où a aussi lieu l'ensemble des processus métaboliques, intègre cette information pour la transmettre à l'axone, "l'antenne émettrice" du neurone. Certains sont recouverts d'une gaine, la *myéline* qui permet d'accélérer la propagation d'une moyenne de $1\,{}^{\mathrm{m}}/_{\mathrm{s}}$ dans le corps cellulaire jusqu'à environ $100\,{}^{\mathrm{m}}/_{\mathrm{s}}$ pour certains axones myélinisés.

Les neurones forment un réseau en établissant des contacts entre eux par des *synapses* qui sont des zones de contact spécialisées assurant le transfert des signaux inter-neuronaux. À titre d'illustration, notons qu'on estime que le SNC humain comporte 10^{11} neurones interconnectés par un total d'environ 240.10^{12} synapses [Koch, 1998, p.87], soit environ 10000 synapses par neurones en moyenne. Celles-ci existent sous forme électrique (le contact est alors direct par l'intermédiaire des canaux ioniques), chimique (par l'intermédiaire d'un messager chimique, le neurotransmetteur) ou encore mixte. Les synapses chimiques sont à elles seules des systèmes dynamiques non-linéaires complexes. Elles sont spécifiques à leur *neurotransmetteur* et existent sous forme excitatrice (85% en nombre) et inhibitrice et s'adaptent aux flux d'activités entre les neurones selon des règles encore méconnues. Généralement, on considère qu'elles réagissent suivant un *conditionnement* : c'est la règle

*à la différence avec un cable électrique où le signal est porté par des électrons libres

de Hebb [1949].

> Quand l'axone d'une cellule A est assez proche pour exciter une cellule B et qu'elle cause de façon répétée et persistante son activation, des mécanismes de croissance ou métaboliques prennent place dans une ou les deux cellules de telle façon que l'efficacité de A, comme une cellule activant B est augmentée.

Cette règle peut aussi s'appliquer de façon contre-associative, c'est alors une règle *anti-hebbienne*.

L'impulsion neuronale : un protocole universel de communication ?

Étudions plus précisément l'électrophysiologie de la création et de la transmission de l'information neuronale. En fait, l'information est intégrée au soma jusqu'à ce qu'un processus non-linéaire "explosif" soit déclenché lorsque le potentiel du neurone atteint un certain seuil : un afflux massif d'ions crée alors un bref et intense pic de potentiel positif, le *potentiel d'action* (PA) suivi d'une période de relaxation où le neurone est réfractaire à toute stimulation[†]. Le PA est activé d'autant plus vite et fréquemment que l'activité excitatrice sur le neurone est forte. Le PA tend à être similaire en amplitude, forme et durée dans tout le SNC : il correspond donc en théorie un signal localisé et binaire. Comme les *bits* des ordinateurs, son caractère binaire est "tout-ou-rien" : l'information neuronale est transmise de façon robuste par l'axone. L'information est donc codée temporellement par un ensemble d'impulsions binaires.

En effet, les PAs sont transmis sans atténuation par les axones jusqu'aux terminaisons synaptiques. Ces dernières s'activent et, en particulier pour les synapses chimiques, libèrent des vésicules de neurotransmetteurs dans l'interstice synaptique. Ce signal génère avec une certaine probabilité au niveau post-synaptique un courant et donc un potentiel électrique, le Potentiel Post-Synaptique (PPS) qui se propage alors jusqu'au soma pour recommencer le processus d'intégration. De façon formelle, on simplifie souvent ce mécanisme

[†]un PA nominal à température corporelle a pour constantes de temps initiation 75 μs, montée, 12 μs et descente, 250 μs

en considérant les PPS comme la réponse impulsionnelle (donc du PA) d'un filtre linéaire (la synapse). Et si l'amplitude des PPS varie d'une synapse à l'autre, on observe par des expériences neurophysiologiques que leur forme est relativement constante pour un neurone.
Plus précisément, les PPS participent typiquement de 3 à 10% à la différence de potentiel pour atteindre le seuil. Aussi, même si l'ensemble de ces processus n'est qu'une simplification de la complexité des comportements neuronaux, il est souvent admis que la contribution des PPS s'ajoute linéairement au soma avec un coefficient de proportionnalité qui caractérise la polarité et la force de la synapse, l'*efficacité synaptique*. Dans cette section, guidés par l'architecture du neurone, nous avons défini de façon implicite un sens dans la direction de la propagation depuis les dendrites vers le soma puis vers l'axone. Toutefois, des mécanismes de propagation des PAs du soma vers les dendrites ont été mis en évidence par Stuart et Sakmann [1994] (propagation antidromique) et semblent corrélés à des mécanismes de contrôle du gain et d'apprentissage. Finalement, ces observations biologiques permettent de définir un modèle complet de réseau de neurones comme un système dynamique complexe.

Capacités computationnelles des neurones

Mais malgré cette complexité des processus neuronaux, le substrat de son information, le PA, est très simple. Aussi, les contraintes physiologiques, comme la fréquence maximale de décharge ($\sim$ 200 Hz) due à la période réfractaire, limitent la capacité computationnelle d'un neurone. De plus, on observe dans le SNC que les neurones sont toujours sous tension, et qu'aucune aire n'est totalement silencieuse, de telle façon que les neurones "tirent" à une fréquence moyenne de 1-10 Hz et que le "son" des neurones[‡] semble être relativement uniforme à travers le SNC [Softky et Koch, 1993b]. Cette observation est sûrement corrélée à la plasticité neuronale, et un neurone ne peut donc pas être considéré comme un simple processus solitaire et statique de calcul. Ils suggèrent l'importance de règles homéostatiques qui contraignent

[‡]C'est-à-dire le son produit par des haut-parleurs lorsqu'on amplifie le signal impulsionnel d'un neurone.

alors les degrés de liberté d'un neurone et donc sa capacité computationnelle théorique.

Comme l'a suggéré Hebb [1949], les neurones, appuyés par les cellules de soutien, créent un code complexe en interagissant pour former des *assemblées*, changeantes au cours du temps et donc l'efficacité est supérieure à la somme de ses sous-parties. De fait, la capacité d'une assemblée conduit à une explosion combinatoire. On peut évaluer théoriquement l'information d'un message par sa *surprise* (mathématiquement, le logarithme base 2 de l'inverse de la probabilité d'un message donne l'information en *bit* [Shannon, 1948]), et donc si la sortie d'un neurone seul est relativement peu "surprenante", la situation est différente si l'on considère le motif spatio-temporel des PAs d'une assemblée de neurones[§].

Dans un premier temps, ceci explique que théoriquement un ordinateur cadencé à 2 GHz sur un mot de 32 bits soit à peine plus efficace que par exemple l'équivalent parallèle et synchrone du cerveau d'une pieuvre dont chacun des 300 millions de neurones produirait un maximum de 200 $^{\text{bits}}/_{\text{s}}$ soit théoriquement, même si cette vision holistique est idéalisée et ne sert bien sûr que d'illustration, 60 GOps[¶]. Ce calcul peut être étendu à l'homme en considérant les 10^{15} synapses qui reçoivent une moyenne de 10 impulsions par seconde. Aussi la structure asynchrone du code le rend autrement plus performant. En effet, dans ce cas, les signaux vont dépendre du décours temporel des impulsions et cette explosion combinatoire est théoriquement encore d'autant plus amplifiée que la précision des impulsions est grande.

En résumé, nous pouvons en déduire que le corrélât du code neural est porté de neurone en neurone par des signaux élémentaires, impulsionnels et robustes, qui se composent pour former le code neural. Cette composition per-

[§]Un argument célèbre examine un singe un peu particulier, entraîné à frapper de façon aléatoire sur les touches d'une machine à écrire comportant pour cette expérience les $M = 256$ caractères ASCII. Chaque caractère porte peu d'information en lui même ($\log_2(M) = 8$ bit) mais l'ensemble aléatoire des caractères tapés peut former un texte, "Les ruines circulaires" de Borgès ou encore le texte de cette page. Aussi, comme malgré tout cet événement (que le texte corresponde exactement) reste très improbable, il porte donc beaucoup plus d'information (plus exactement $N.\log_2(M)$ bit où N le nombre de lettres du texte produit.

[¶]soit 60 milliards d'opérations par seconde

met alors de créer grâce à une distributivité massive et au caractère temporel et asynchrone du codage impulsionnel un potentiel calculatoire énorme.

2.1.3 Flux d'information dans l'architecture du SNC

Ce dernier point suggère l'importance de l'asynchronie et du codage temporel dans l'efficacité du SNC. Nous allons voir, en étudiant le cas particulier de l'organisation corticale, comment ces principes s'intègrent à différents niveaux dans le SNC pour fournir ensuite les bases d'un modèle du code visuel.

Le calcul cortical

L'importance de la communication entre neurones se reflète dans leur organisation. En effet, on observe que les neurones sont distribués dans le SNC de façon non-uniforme. Ainsi, le développement du cerveau chez l'embryon suggère que l'Évolution a préféré adopter une structure laminaire (dite de dimension $2+\varepsilon$) qui privéligie des organisations suivant des surfaces. En particulier, cette organisation en *couches neurales* est remarquable dans le cortex (du grec *écorce*), c'est-à-dire la partie la plus externe du cerveau. Le cortex est supposé être apparu le plus tard dans l'Évolution autour d'un noyau commun constitué de l'épine dorsale et du mésencéphale. C'est une surface dense en neurones, richement plissée (chez l'homme, le cortex déployé couvre environ $1\,\mathrm{m}^2$) qui malgré sa diversité fonctionnelle développe sur toute son étendue une structure similaire. Il est constitué de 6 sous-couches de neurones (numérotées depuis l'extérieur I à VI et dont chacune contient des cellules pyramidales) selon un principe qui se répète tangentiellement à sa surface : une couche granulaire (la couche IV, ainsi appelée pour la forme des neurones qui la constituent) qui le divise alors en couches supra-granulaires sur le coté externe (les couches I à III) et les couches infra-granulaires sur le coté interne. Le cortex est formé d'une majorité de neurones pyramidaux d'orientation radiaire et dont leur apex dendritique est dans la couche I.
Les axones des afférents du thalamus traverse la couche infra-granulaire et forment des synapses (exclusivement) dans la couche IV avec les dendrites des neurones des autres couches. L'information est alors propagée suivant

un chemin axial, définissant ainsi une *colonne corticale* d'environ quelques centaines de neurones, et qui peut définir un automate indépendant structurant un réseau cortical adaptatif [Burnod, 1989]. Les neurones de la couche supra-granulaire projettent alors l'information par des relations symétriques (c'est-à-dire réciproques) sur les neurones d'autres colonnes proches (pour les neurones de la couche III) ou plus distaux (pour ceux de la couche II). Les neurones des couches infra-granulaires reçoivent cette information (de façon non-réciproque) et peuvent alors la retro-propager au thalamus, créant ainsi une boucle dans le flux de l'information.
Le flux d'information arrive donc en premier par une voie "en avant" où l'information en provenance d'autres régions du SNC (par exemple des entrées thalamiques) est intégrée, via la couche granulaire, par la couche *supra*. Une voie latérale propage alors cette information aux colonnes voisines. Une autre voie "descendante" va intégrer l'activité corticale locale sur les couches *infra* pour la renvoyer à la source. Une colonne peut être partiellement (seulement les couches supra) ou globalement activée par l'excitation simultanée des cellules des couches infra et supra. Une activation globale peut induire une action (motrice par exemple) par l'intermédiaires des axones des neurones des couches supra-granulaires. Une activation partielle (donc de la couche supra) va donc plutôt correspondre à un état d'anticipation, "attendant" l'activation de la couche infra. Cette "boucle", en renvoyant un signal sur les sources, permet notamment de réguler l'activité corticale. Par analogie avec les protocoles de calcul asynchrone, on peut interpréter ce signal de retour comme un signal de *hand-shake*, soit un "accusé de réception" de l'événement cortical.

Les aires corticales

En élargissant encore plus notre point de vue, on observe que la surface du cortex est découpée en aires corticales, et qu'il semble apparaître des séparations physiologiques qui permettent de définir une *carte corticale*, comme par exemple la classification en aires de Broadman. Toutefois, les limites de ces aires ne sont pas toujours précises, dépendent de leur définition

(par le type de cellule ou par la morphologie corticale par exemple) et ne correspondent pas forcément à des séparations fonctionnelles. Cette imprécision est confirmée par la plasticité du SNC qui peut modifier la répartition des aires corticales. On peut par exemple déterminer chez le macaque des sous-aires corticales dans le cortex moteur correspondant aux doigts qui projettent des nerfs moteurs vers la main. On peut alors montrer qu'une invalidité permanente d'un doigt (qui par exemple a été sectionné) va modifier les sous-aires en réorganisant la place de la sous-aire du doigt non-utilisé.
Cette structure va favoriser l'activation parallèle et distribuée des processus, suivant le modèle de McClelland et al. [1986]. En effet, elle va favoriser les associations grâce au riche réseau de connexions cortico-corticales entre aires. On remarque que de façon générale afin de minimiser le volume des connexions, si deux fonctions corticales sont similaires ou corrélées, alors leurs localisations seront voisines. Ce genre de processus, implanté pour des réseaux de neurones artificiels par Kohonen [1982], permet de construire des *cartes topologiques* à partir des statistiques des entrées en minimisant l'erreur de représentation de l'ensemble des entrées sur la surface de la carte. Cette méthode permet par exemple de projeter sur une surface des signaux de grande dimension comme les similarités entre des visages, et donc de placer côte-à-côte sur une carte les visages similaires. Ainsi, comme il a été observé dans le gyrus fusiforme, une aire corticale pourrait représenter l'ensemble des formes du visage sur un *espace de visages* [Leopold et al., 2002]. À l'inverse, on pourra alors définir un visage intermédiaire comme la projection inverse de l'intermédiaire topologique (sur la surface corticale) des deux représentations. Les interactions entre colonnes, qui peuvent alors se regrouper pour former des *hypercolonnes*, comme une assemblée de colonnes, peuvent alors créer des modularités nouvelles, et on peut imaginer selon le même principe qu'alors une hypercolonne pourrait différentier suivant des surfaces différentes et imbriquées les différentes expressions du visage.
Il faut alors noter l'importance des interactions cortico-corticales distales. En effet, les efférences de la couche II peuvent s'organiser en fibres, les fibres associatives, qui se dirigent vers d'autres aires corticales de façon très précise, notamment pour rentrer en contact avec des régions de l'hémisphère contra-

latéral. En particulier, on rencontre souvent des régions corticales connectées réciproquement se projetant sur de mêmes aires tierces. Pour synthétiser, si l'on image le flux de l'information depuis une entrée thalamique, alors cette information va se propager en parallèle sur la surface corticale pour activer certaines colonnes. Ces colonnes, en plus de renvoyer au thalamus un signal, vont propager cette activité à d'autres aires corticales. Comme on a vu que la réponse sera traduite temporellement (en général d'autant plus rapidement que l'activité est forte) et que l'activité d'une colonne dépendra aussi de son état interne, le flux de l'information dans le cortex va donc obéir à un processus parallèle et distribué qui va être influencé par le décours temporel de la propagation de l'information.

Architecture globale du flux d'information

> The problem that continually tends to be overlooked is that of the formation of those "primary units" to which the inferential activity of the cognitive process can apply. [...] The primary process cannot be considered a passive, mechanical recording of distant stimuli, but consists of an organization of proximal stimulation [...] transforming an unrelated set of elements [...] into a certain number of segregated units with precise spatial and temporal relationships.
>
> Kanizsa [1979]

S'il est difficile d'enregistrer l'activité simultanée des neurones d'une colonne, des techniques d'imagerie cérébrale comme l'Imagerie par Résonance Magnétique (*IRM*) ou les potentiels évoqués mesurés par l'Électro-EncéphaloGramme (EEG) permettent de localiser dans l'espace et le temps de façon relativement satisfaisante les aires du cortex activées par une stimulation. On peut alors faire correspondre des fonctions à des localisations corticales afin d'en étudier les principes neurophysiologiques [Logothetis et al., 2001b]. Ces techniques d'imagerie permettent de contraster des activités cognitives et confortent l'idée que des fonctions différentes sont effectuées dans des aires différentes. Aussi, ces techniques d'imagerie combinées avec des techniques poussées de

traitement du signal [Delorme et al., 2001a] ont montré que l'activation des aires se fait de manière précise et asynchrone à travers le SNC comme une succession d'événements corticaux.
En considérant le SNC du primate en entier maintenant, on en déduit des principes fondateurs de son organisation. Dans l'optique de modéliser ce système, on peut être tenté de simplifier dans un premier temps son organisation en regroupant l'ensemble des réseaux autour d'un noyau central (qui correspondrait au système mésencéphalique plus ancien au sens de l'Évolution, comme en particulier le thalamus), spécialisé dans la réception des signaux sensoriels et de son "routage" vers le cortex, contrôlé par les boucles de retour. Le flux d'information va se propager comme une succession d'événements corticaux se propageant dans l'ensemble du SNC. Ce flux d'information est organisé à plusieurs niveaux par associations pour produire par exemple à partir du cortex moteur les actions musculaires.
L'organisation du SNC montre donc clairement que la propagation de l'activité cérébrale est dynamique du niveau du neurone jusqu'au niveau cortical. Le corrélât neuronal du code visuel est donc organisé suivant la coopération entre la propagation de l'information visuelle et les transformations temporelles de cette information dans les aires corticales, où la confrontation entre des processus de reconnaissance et des processus locaux de coopération ou de compétition induisent la mise en avant de certains aspects, *saillants**, de l'information. Le flux temporel est organisé suivant une voie en avant mais est ensuite influencé par le retour de l'information des couches infra. Finalement, les processus neuronaux, comme la Vision, ne sont donc pas une suite de calculs, mais plutôt un flux d'information précisément transformé par les interactions entre les neurones, une suite d'événements dans une structure adaptative. Afin de mieux saisir le fonctionnement particulier de ces interactions dans les processus visuels, nous allons nous consacrer maintenant plus spécifiquement au système visuel.

*On utilisera dans la suite, par analogie avec l'anglais, le substantif *saillance*.

2.2 Anatomie du système visuel

Les principes généraux du SNC que nous avons montré vont se matérialiser d'une façon spécifique pour le système visuel. Aussi nous allons définir précisément l'anatomie fonctionnelle du système visuel en suivant l'information lumineuse au cours de son trajet depuis l'œil jusqu'à la réponse motrice.

2.2.1 La rétine : de l'information lumineuse au codage impulsionnel

Tout d'abord, la *rétine* est une partie très particulière du SNC : cette partie du cerveau, dont la surface peut être visible simplement par rétinoscopie, s'isole du reste de la masse cérébrale tôt dans le développement de l'embryon. Sa morphologie et son fonctionnement complexes sont donc plus facilement observables que le cortex . La rétine est donc un morceau de choix pour initier l'étude du système visuel.

Morphologie de la rétine : de la lumière à l'impulsion

La rétine est une mince surface d'environ .5 mm située au fond de chaque œil, couvrant $\sim 75\,\%$ du globe oculaire. Elle constitue la partie sensitive de la vision en transformant l'image lumineuse focalisée par l'œil en un signal de PAs. Comme le reste du SNC, elle est constituée de neurones accompagnés par des cellules de soutien et est largement innervée par des vaisseaux sanguins. Suivant les observations de Cajal [1911], dont les coupes étaient souvent préparées verticalement à la surface, on sait que la rétine est constituée de 6 couches successives de neurones différentiés, depuis la couche des neurones sensibles à la lumière, les *photorécepteurs* (PhR) jusqu'à sa sortie, la couche des *cellules ganglionnaires* (CG), dont les axones forment le nerf optique.

Les PhRs sont les neurones sensibles à la lumière et peuvent grâce à leur capacité d'adaptation répondre à de très faibles intensités (dans des conditions expérimentales précises, un photon seul peut suffire à les stimuler). Ils existent sous deux formes : les cônes (à réponses rapides et sensibles sui-

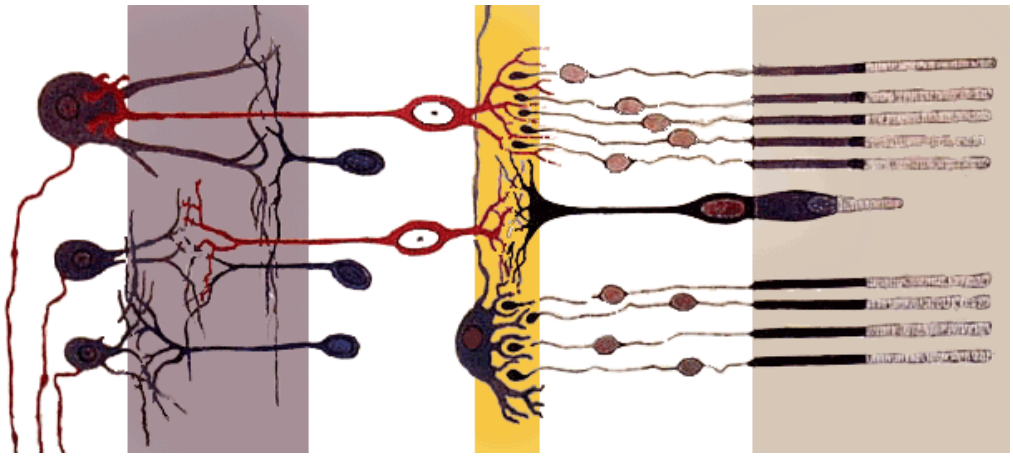

FIG. 2.2.1 – *Organisation axiale simplifiée de la rétine.* La lumière focalisée par l'œil traverse (ici de gauche à droite) les couches rétiniennes pour activer les photo-récepteurs (qui se présentent sous forme de cônes ou de bâtonnets, couche à l'extremité droite). Ceux-ci activent en avant (axialement, ici vers la gauche) les cellules bipolaires et latéralement par les cellules horizontales, de façon à ce que le signal soit sensible aux contrastes spatiaux. Le signal se propage en avant vers les Cellules ganglionnaires (CG, dont les corps cellulaires forment la dernière couche, ici à gauche), dont on voit apparaître deux tailles correspondant aux voies Magno et Parvo. Ce signal est filtré latéralement par les cellules amacrines, qui elles sont principalement sensibles aux contrastes temporels. (modifié depuis un dessin de Cajal)

vant leur type à différentes gammes de longueur d'ondes de la lumière) et les bâtonnets (plus lents mais sensibles en vision nocturne) et convertissent par un processus photo-chimique l'énergie lumineuse en un signal neurochimique (ils ont la particularité physiologique d'être continuellement activés en l'absence de stimulus). La réponse analogique des PhRs à la puissance lumineuse suit approximativement une loi logarithmique de telle sorte que la comparaison de différentes réponses mesurent des rapports de puissance lumineuse. Ils diffusent cette activité grâce à leurs synapses avec les cellules bipolaires et horizontales, de telle sorte que les activités de ces cellules sur la surface rétinienne forment une image des contrastes lumineux. Ce signal est ensuite propagé en avant des cellules bipolaires jusqu'aux CGs. Les cellules horizontales et amacrines jouent alors un rôle important en propageant en même temps le signal latéralement dans la rétine.

Étudions plus spécifiquement la nature des signaux rétiniens. L'image rétinienne, image optiquement distordue et inversée de l'information lumineuse, est transformée analogiquement à travers la rétine en une image complexe correspondant à l'activité des neurones des différentes couches. En effet, depuis le signal synaptique émis par les PhRs, l'information est codée par un signal analogique se propageant par des contacts synaptiques et par diffusion chimique. Les cellules bipolaires sont sensibles à des contrastes spatiaux alors que les cellules amacrines sont plus particulièrement sensibles à des variations temporelles. Elles vont ainsi transformer temporellement l'image des contrastes lumineux qui vont être propagé en avant vers la couche des CGs. Toutefois, ces comportements sont plus complexes et variés et correspondent à toute une famille de transformées spatio-temporelles qui ne sont pas encore aujourd'hui toutes connues (pour une revue, on pourra consulter [Meister et Berry II, 1999]). Finalement, c'est seulement à la couche des CGs (donc la plus interne) que les contrastes lumineux peuvent produire des impulsions neuronales qui constitueront alors entièrement le signal impulsionnel emprunté par le nerf optique. La rétine a alors achevé la transformation de l'information visuelle en un train d'impulsions.

Organisation spatiale des neurones de la rétine : de l'image lumineuse à l'image rétinienne

Cette approche radiale de la rétine a été affinée par des techniques instrumentales modernes qui ont montré l'importance de l'organisation spatiale sur la surface de la rétine. Comme Cajal l'avait déjà observé, la concentration et la nature des différents neurones varient suivant la position sur la rétine. En effet, on observe tout d'abord que la concentration en cônes augmente fortement quand on s'approche de l'axe de vision de l'œil, le *fovéa* (il n'y a presque plus de cônes en dehors du disque d'un degré depuis le fovéa*, la *macula*). En complément, la concentration en bâtonnets y est nulle mais atteint un maximum vers 3 deg. Cette observation explique que nous n'avons pas de sensibilité à la couleur hors de la fovéa†, et de façon symétrique que pour voir une étoile faible, il est conseillé de fixer son regard légèrement à coté pour que la lumière de l'étoile frappe les bâtonnets plutôt que les cônes. On définit alors pour les cellules ganglionnaires leur *champ récepteur* [Imbert, 1983] comme l'ensemble des PhR qui participent à son excitation‡. On observe alors que ceux-ci ont des contours sensiblement circulaires dont le rayon grandit en moyenne proportionnellement à leur excentricité. Depuis le fovéa —où il équivaut à un PhR (cône) par CG— les champs récepteurs peuvent atteindre 10 deg de champ visuel à la périphérie de la rétine. On peut donc définir une *carte rétinotopique* qui va correspondre à la transformation spatiale de l'image par cet arrangement spatial de la grille des neurones. Elle est souvent approchée par une transformée log-polaire de l'image sphérique centrée sur l'œil§. Cette carte est incomplète dû au recouvrement de régions par les axones des CGs (le scotome) et notamment à la conver-

*Soit environ la taille d'un ongle à bout de bras.

†Cette observation contre-intuitive peut être vérifiée par un dispositif expérimental simple : on fixe la direction du regard et on présente des stylos de couleur en périphérie (à environ 10 deg) : les couleurs sont perçues illusoirement, mais ne peuvent être déterminées avec précision.

‡On définira par extension les champs récepteurs des neurones dans le système visuel de la même façon.

§Il est intéressant de noter que si la concentration de PhRs baisse proportionnellement à l'excentricité et que l'image d'un cercle centré sur le fovéa touchera un même nombre sensiblement égal de PhRs, ce qui introduit une stabilité *par structure*.

gence des axones vers le nerf optique, qui constituent une zone insensible, le *point aveugle*.
Finalement, les CGs sont sensibles à des contrastes de lumières à plusieurs tailles, et qualitativement certaines CG répondant maximalement lorsque le signal correspond à un spot de lumière entouré d'un pourtour de la taille de son champ récepteur (cellule centre-ON) ou son inverse (centre sombre sur fond clair, les cellules centre-OFF). Aussi, même si les réponses des cellules sont très variées et malgré la complexité du réseau rétinien, Rodieck [1965] a montré que la réponse jusqu'au soma des CGs pouvaient être modélisée de façon linéaire par rapport aux réponses des photorécepteurs. Cette simplification peut théoriquement permettre de déterminer de façon exacte la transformée d'une CG (qui est alors un filtre spatio-temporel linéaire) grâce à sa réponse impulsionnelle à un spot de lumière. Toutefois, de nombreux phénomènes non-linéaires sont introduits dans les réponses rétiniennes comme la perception de la couleur par des mécanismes non-linéaires [Alleysson et al., 2002]. De plus Sherman et Guillery [1996] a montré que le contexte pouvait modifier l'activité des CGs.

Transformation multicanaux : l'image impulsionnelle

On a vu que ce sont seulement les CGs qui émettent les PAs qui seront transmis au reste du SNC, ce qui montre que depuis les 10^8 photorécepteurs (PhR) via environ 10^9 cellules intermédiaires et jusqu'au nerf optique constituée par les axones du million de cellules ganglionnaires, la transformée rétinienne est une transformation d'une intensité lumineuse variant dans le temps à un signal spatiotemporel impulsionnel de PA. On remarque aussi que le nombre relativement peu élevé de fibres en sortie montre que la taille du signal doit être comprimée pour qu'il soit transmis efficacement au reste du SNC. Une méthode "utilisée" par la rétine est alors de transformer l'information visuelle en un signal multicanaux qui tend à séparer les sources qui ont produit la sensation lumineuse, réduisant ainsi la dimension du signal à transmettre[¶].

[¶]Mathématiquement, ce mécanisme se rapproche donc d'un transformée en composantes orthogonales.

En particulier, Atick et Redlich [1992] ont montré que la réponse des cellules ganglionnaires à différentes fréquences spatiales coïncidait avec une réduction des corrélations spatiales entre des locations voisines, montrant ainsi que des principes écologiques peuvent guider la compréhension des fonctions rétiniennes. Un aspect du codage rétinien est donc de souligner des parties non redondantes et qui sont donc relativement saillantes. Dans notre cadre, cette sensibilité va permettre de propager plus rapidement les parties de l'image les plus saillantes, conduisant de plus à une transformation temporelle de l'information spatiale. De façon analogue, on observe que les cellules ganglionnaires transforment l'information lumineuse en signaux relativement indépendants. Ainsi l'information de couleur, la *chrominance* est séparée de l'information d'intensité lumineuse, conduisant à un "multiplexage" de l'information lumineuse [Roska et Werblin, 2001]. On observe ainsi que des CGs morphologiquement et fonctionnellement différentes (cellules α, β et γ) vont porter des canaux différents. Ce découplage sera aussi temporel puisque l'information d'intensité lumineuse est plus rapidement activée que la couleur, créant ainsi des voies à plusieurs latences pour l'information rétinienne.

Finalement, si on présente assez rapidement une image à un sujet pour éviter toute saccade oculaire, celle-ci va se projeter au fond de l'œil en une image distordue et inversée, activer les photo-récepteurs puis tout le réseau rétinien pour enfin être transformée en de multiples canaux par les cellules ganglionnaires. De manière synthétique, chacune de ces cellules peut alors être caractérisée par une sensibilité maximale à un canal particulier et par une réponse temporelle, mais les sensibilités peuvent se recouvrir avec celles d'autres CGs et sont interdépendantes [Salinas et Sejnowski, 2001]. L'image que nous percevons est alors entièrement codée en un train d'impulsions en environ $20 - 40\,\text{ms}$ (voir Fig. 2.2.2-(A)). Alors que la vague d'activité rejoint maintenant le nerf optique, le décodage de cette transformation dans le reste du système visuel semble alors tenir du miracle.

2.2.2 Propagation de l'image vers les aires visuelles

Nous allons maintenant nous concentrer sur la propagation du flux de PAs que nous avons initié et suivre cette vague d'activité jusqu'aux aires supérieures (voir Fig. 2.2.2).

Du Corps Genouillé Latéral (CGL) à l'aire visuelle primaire (V1)

Les fibres du nerf optique se séparent au chiasma (la décussation, voir Fig. 2.2.2-(B)) de telle sorte que chaque hémichamp de l'image depuis l'axe vertical (gauche ou droit) soit transmis à un hémisphère cérébral ipsilatéral (*i.e.* chaque champ gauche rejoint l'hémisphère gauche). Le signal maintenant stéréoscopique converge alors au *Corps Genouilé Latéral* (CGL, voir Fig. 2.2.2-(C)), une formation thalamique composé de 6 couches, au bout d'environ $30 - 50$ ms. En regroupant les axones des différents types de CGs, la structure laminaire sépare distinctement les voies *Magno* de la voie *Parvo**, et donc les réponses sont multiplexées. Toutefois le CGL ne semble pas chez le primate opérer de transformation, mais son importance fonctionnelle suggère que son rôle de relais à une forte importance computationnelle. Ainsi, il semble que des neurones du CGL puissent être bloqués suivant le contexte, ce qui expliquerait alors les processus d'interruption ou de sélection de certains afférents qui peuvent être utilisés dans certains mécanismes attentifs. Il pourrait aussi être à l'origine de la modification de la transmission de l'information lumineuse lors de saccades†, ou encore lors de rivalités binoculaires [Logothetis et al., 2001a]. Il est important de noter que les fibres correspondant au fovéa divergent principalement vers l'aire 7 (suivant la classification de Broadman) via une autre formation thalamique, le tectum.
Depuis le CGL, la vague d'activité atteint alors le cortex par l'*aire visuelle primaire* (V1, l'aire 17 dans la classification de Broadman, voir Fig. 2.2.2-(V1)) après environ $40 - 60$ ms‡. Cette aire est beaucoup plus vaste que

*Nous ignorerons ici la voie *Konio*, encore trop peu connue.

†On parle souvent abusivement de suppression, mais il a été prouvé qu'un perception est possible durant la saccade [Castet et Masson, 2000]

‡Elle atteint aussi l'aire MT, mais cette voie semble privilégier l'information de mouvement et nous ne la traiterons donc pas dans le cadre de notre étude.

la rétine et le CGL puisqu'à chaque fibre entrante va correspondre plusieurs centaines de neurones (V1 totalise chez l'homme environ 350 millions de neurones). V1 est une aire corticale organisée verticalement, comme nous l'avons décrit dans la Sec. 2.1.3, et les travaux de Hubel et Wiesel [1959, 1962] ont montré qu'elle était formée d'hyper-colonnes montrant une sélectivité à la direction du contour d'un stimulus visuel. Des méthodes d'imagerie [Tanaka, 1996] permettent de mettre en évidence la carte de la répartition de ces orientations, montrant une inter-digitation des différents stimuli. Cette organisation est caractéristique de la projection d'un signal multidimensionnel sur la surface bidimensionnelle du cortex par des processus locaux de compétition et coopération, et forme des bandes dites de Turing [Turing, 1952].
Plus précisément, les fibres de la voie Magno, correspondant à l'information rapide de luminance, projettent sur une sous-couche de la couche corticale granulaire, la couche IV-C_α de V1 pour former une image (principalement stéréoscopique et des contours) puis contactent comme nous l'avons décrit les couches supra-granulaires II et III (les cellules simples). Une lésion de ces aires entraîne une disparition totale de toute perception visuelle correspondant à la surface lésée. La propagation peut alors être modifiée par les neurones de la couche IV-B qui sont spécifiquement sélectifs à des mouvements dans la direction perpendiculaire à la direction préférée (les cellules complexes) et ainsi "prévoir" la réponse future. Ainsi le contexte d'orientation, direction du regard, profondeur sont autant de facteurs qui influencent la réponse des neurones de V1, et soulignent l'importance du contexte dans les processus cognitifs.

L'aire visuelle secondaire (V2)

Les neurones de V1 projettent ensuite sur l'aire corticale voisine *V2* après un total d'environ 50-70 ms. La fonction de cette aire, de surface légèrement plus réduite, semble être semblable à première vue, les neurones étant sensibles à des barres à différentes orientations mais sur des champs récepteurs plus larges [Wallis et Rolls, 1997], un phénomène d'élargissement qui se généralise en allant vers les aires supérieures. Toutefois, Leventhal et al.

[1998] a prouvé que la réponse des neurones de V2 pouvait être indépendante de la façon dont ces barres sont dessinées, c'est-à-dire qu'un même neurone répondra à une barre pleine et à la même barre dessinée avec une texture.
Ce résultat suggère un principe général d'organisation corticale. Tout d'abord il montrerait qu'à la suite de la transformation dans la rétine, un des principes des processus neuronaux pourrait être de décomposer le signal en canaux "indépendants", en séparant ici le contour de la texture. Ceci revient à un processus d'ingénierie inverse dans le sens où le signal est transformé depuis sa perception pour essayer de séparer des processus causaux distincts qui l'ont généré, comme ici la forme d'un objet va être déduite de son contour alors que la nature de sa surface va l'être de sa texture. Aussi, dans notre cadre, ce résultat montre que le code visuel va pouvoir traduire ce résultat en un signal spatio-temporel et que le flux d'information va à ce niveau diverger entre des informations parallèles et autonomes.

Les aires supérieures

Après V1 et V2, le flux d'information se distribue dans plus de 30 aires. Ces aires sont elles-mêmes largement inter-connectées dans un réseau complexe mais on peut toutefois extraire de ce réseau d'aires corticales deux voies principales, la voie dorsale, qui se concentre sur les mouvements et la localisation des objets (voie "Où ?") et la voie pariétale, qui se concentre sur l'identification de la scène (voie "Quoi ?") [Ungerleider et Desimone, 1986] et qui forment ainsi les *chemins de la vision*. Plus particulièrement le long de la voie "Quoi" le flux atteint V4, puis les aires Postérieure Inféro-Temporale (PIT), Centrale Inféro-Temporale (CIT) et finalement Antérieure Inféro-Temporale (AIT) après environ $80 - 100$ ms (voir Fig. 2.2.2).
En général, les champs récepteurs des neurones grandissent et sont par exemple cent fois plus grands dans CIT que dans V1, de telle sorte que la rétinotopie s'estompe et qu'en complément, ces neurones deviennent de plus en plus sélectifs à des objets précis de façon de moins en moins localisée, jusqu'à certains neurones corticaux recouvrent le champ visuel entier. L'image est donc transformée au profit d'une représentation progressivement plus fonction-

nelle. On peut dégager depuis des études neurophysiologiques des fonctions spécifiques à ces aires, les neurones de V4 sont particulièrement sensibles à l'orientation et ceux de V3 aux formes, laissant suggérer que l'image serait décomposée en attributs de plus en plus complexes abstraits [Zeki et Shipp, 1998] mais les comportements non-linéaires de neurones répondant à des signaux complexes montrent que cette théorie est limitée et que ces ségrégations ne sont pas strictes.
Les neurones du cortex inféro-temporal (IT) sont alors particulièrement pertinents dans le cadre de notre étude car une lésion de ces aires affecte spécifiquement la reconnaissance des objets mais sans en perdre la perception. Plus précisément, de nombreuses expériences d'enregistrement ont été conduites sur des neurones de cette région en présentant de larges variétés d'objets pour en étudier la sélectivité. Ainsi, des neurones de IT sont spécifiquement sensibles aux mains, au yeux, ou à d'autres objets naturels ou artificiels [Tanaka, 1996; Vogels, 1999]. Aussi, cette sensibilité est souvent indépendante du mode de représentation (photo ou dessin) alors qu'elle peut dépendre énormément du contexte. Enfin une exposition prolongée à des objets ou des images modifie la sensibilité de ces neurones. En simplifiant ce modèle et s'il existe bien des processus adaptatifs similaires laissant émerger des propriétés statistiquement indépendantes, on comprend alors comment peuvent alors apparaître spontanément des neurones aussi spécifiques que ceux sensibles à l'image de Bill Clinton que nous avons évoqués dans la Sec. 1.2.3.

2.2.3 Représentations temporelles de l'information visuelle

En cheminant le long de l'architecture du système visuel nous avons vu la multiplicité et la complexité de son organisation et de son code par rapport à l'unité que l'introspection nous suggère. Pour aborder le code de la Vision, nous allons donc essayer de comprendre quelle forme prend la représentation de l'information visuelle aux différents niveaux de son observation.

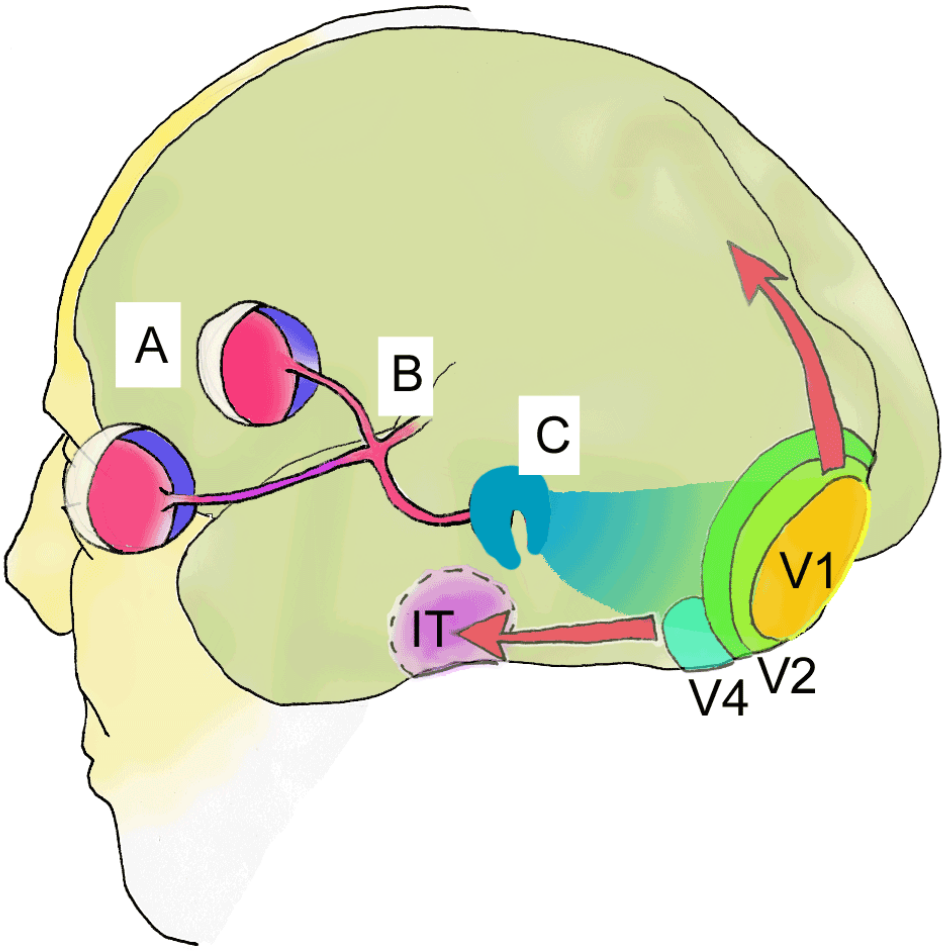

FIG. 2.2.2 – *Les Chemins de la vision.* Cette figure synthétise schématiquement l'organisation des premiers modules du système visuel impliqués dans la catégorisation ultra-rapide. Depuis (**A**) les yeux, la rétine forme le nerf optique qui se sépare au chiasma (**B**) formant la commissure optique jusqu'au tectum (non représenté) et au CGL (**C**). La commissure geniculo-calcerine projette sur V1 (**D**) puis vers les aires visuelles supérieures (**V1**) et (**V4**), jusqu'au cortex inféro-temporal (**IT**) . Nous avons aussi indiqué schématiquement les voies "Où" et "Quoi" qui montrent la spécialisation hiérarchique des voies visuelles.

Représentations cellulaires

L'organisation que nous avons présentée dans la section précédente, avec une structure majoritairement hiérarchique du système visuel pousse alors à penser que des aires corticales plus élevées vont présenter des sélectivités encore de plus en plus fines. En la caricaturant, cette théorie supposerait alors la présence d'un neurone "grand-mère", sensible alors spécifiquement à la vision de son aïeule (ou comme nous l'avons déjà évoqué à celle de Bill Clinton). Outre le fait que ce type de neurone soit rarement observé en pratique, cette représentation serait vraiment trop coûteuse en termes informationnels. Comme nous l'avions évoqué (voir Sec. 2.1.3), on préfère alors penser que les neurones d'une aire corticale s'organisent localement pour que les sensibilités similaires resteront dans le même voisinage cortical. C'est ce que l'on observe par exemple dans V4 où par exemple on peut trouver des neurones voisins ayant des sensibilités maximales à des cercles concentriques dont les transformations usuelles (translation, agrandissement) seront voisines [Gallant et al., 1996]. Cette représentation permet donc de définir des classes de stimuli (ici des courbes similaires) qui permettent par exemple simplement d'implanter une distance dans une carte corticale (par exemple le temps de propagation entre deux neurones) et d'utiliser la position des neurones comme un outil computationnel.
Ainsi, on peut observer dans IT des neurones sensibles à un objet tridimensionnel simple (e.g. un trombone tordu) et si l'on fait tourner à cet objet, des neurones voisins seront activés : l'objet et ses vues semblent donc être représentées explicitement. On peut alors trouver des neurones voisins qui seront d'autant plus proches que cette rotation est proche [Tanaka, 1996]. Cette expérience neurophysiologique permet d'interpréter les résultats psychologiques de Shepard et Metzler [1970] dans une expérience de manipulation mentale d'objets tridimensionnels pour lesquels le temps de réponse était linéairement proportionnel à l'angle de rotation en suggérant que cette rotation a bien lieu sous la forme d'un déplacement de l'activité sur cette aire corticale, soit donc un déplacement de la représentation. Cette notion étend donc la notion de "mot" neuronal à une carte discrète par la position

des neurones mais continue dans la réponse de son implantation corticale. Surtout, elle permet de traduire par la simple position de l'activité corticale des neurones une transformation complexe comme la réponse d'un neurone à l'image d'un visage dont on fait varier la direction de l'éclairage [Hietanen et al., 1992].
Il est toutefois sûrement incorrect d'avoir une approche seulement symbolique sur la sélectivité d'une carte. En effet il est tentant, surtout pour implanter des réseaux de neurones complets sur un calculateur*, de supposer que les sélectivités des neurones se déduisent entre elles par exemple par de simple transformations géométriques. Même si cela semble vrai dans V1, ce ne l'est plus par exemple dans d'autres aires comme IT, car par exemple des neurones sensibles à des visages dans cette aire peuvent ne pas être activés par la présentation de visages à l'envers [Jeffreys, 1989]. Les cartes corticales n'obéissent donc pas à des règles simples comme des translations ou des rotations mais semblent plutôt émerger des associations répétées dans l'ensemble des stimulations. C'est seulement quand celles-ci obéissent à une certaine généralité (comme celle induite par un mouvement simple dans un environnement) que des règles - comme l'invariance par translation - peuvent apparaître.

Représentations par des assemblées neurales

Au niveau des assemblées de neurones, il semble admis aujourd'hui que la réponse d'un neurone à un flux impulsionnel sera surtout corrélée au degré de synchronie de l'activation des synapses correspondant à la "signature" de ses entrées préférées. Tout d'abord, les expériences de MacLeid et al. [1998] sur le bulbe olfactif montrent que si l'on empêche la synchronisation des entrées tout en préservant la même activité, alors la réponse peut être annulée. De la même manière, Stevens et Zador [1998] ont montré en combinant expériences et simulations que la réponse irrégulière des neurones *in vivo* était le reflet

*La limitation réside alors principalement dans la taille de la mémoire. Pour ne pas stocker tous les paramètres d'un réseau, on est obligé dans un calculateur classique de supposer que les paramètres des neurones se déduisent par les transformations simples depuis quelques prototypes.

de leur préférence pour des entrées synchrones, suggérant alors que la synchronie pouvait être un mécanisme généralisé de communication dans une couche neurale déclenchant l'activité d'un neurone.
La représentation dans une assemblée de neurones est donc créée par la synchronisation des impulsions. En particulier, il a été montré chez les primates que l'engagement dans une fonction cognitive était corrélé avec l'apparition dans la zone corticale correspondante d'oscillations autour de 40 Hz (les ondes γ), interprétées comme la contribution synchronisée d'un large nombre de neurones [Fries et al., 2002], mais quon peut aussi interpréter comme l'artefact .
Ces remarques revendiquent donc un modèle distribué de codage de l'information et suggèrent une représentation par l'activation temporelle simultanée (et souvent rythmique) d'ensembles précis de neurones. Or il est surprenant de voir dans la littérature que l'asynchronie d'un code, c'est-à-dire son décours temporel, est souvent assimilé à l'incohérence de son message. Ainsi, alors que nous avons vu l'importance que jouait le temps dans les processus cognitifs mais aussi l'importance de l'ordre de la propagation du flux d'information dans l'architecture des réseaux de neurones biologiques, l'ordre précis d'arrivée d'impulsions sur un neurone est écarté. En fait, la technologie de mesure neurophysiologique de l'activité des neurones est souvent limité à un petit nombre de neurones, dont on ne peut souvent prévoir ni les caractéristiques ni les inter-relations. On peut donc imaginer que les progrès dans la technologie des multiéléctrodes mais aussi du choix des neurones seront déterminants pour comprendre le code neural dans une assemblée.

Une représentation visuelle globale ?

Toutefois, à un niveau plus global, les travaux de neurophysiologie montrent que les processus cognitifs peuvent être influencé par le décours exact de l'activité et de façon plus générale par le contexte temporel, c'est-à-dire par l'historique de l'activité d'un stimulus ou d'une fonction cognitive. C'est l'aveu d'une grande dépendance des processus cognitifs de l'asynchronie des stimuli au niveau global du SNC et à plus long terme. Il parait alors évident à

ce niveau que si l'on inverse l'ordre des stimuli dans une opération de conditionnement, comme celui du conditionnement du processus de salivation d'un chien au son d'une cloche les résultats seront complètement différents. Cet aspect asymétrique est exploité en particulier dans la théorie du renforcement [Sutton et Barto, 1991].

Aussi, les études de Rodriguez et al. [1999] et Varela et al. [2001] ont montré chez l'homme que l'on pouvait grâce à l'EEG détecter chez l'homme des oscillations γ dans différentes aires visuelles, et que celles-ci pouvaient être bloquées en phase, c'est-à-dire que pendant une oscillation, le décours temporel entre les aires visuelles était constant. Cette observation peut-être interprétée et simulée de façon très simple par un réseau événementiel c'est-à-dire asynchrone, dont la réaction va dépendre de l'histoire de l'activité de ses afférents. La représentation visuelle globale n'est alors que la trace de l'ensemble des activations et peut donc prendre des voies différentes au cours du temps au gré du bruit et de l'adaptation neuronale. Dans notre cadre, nous retiendrons qu'une représentation globale de l'image correspondra au décours de l'activation dans le SNC.

On comprend alors pourquoi nous avions dit qu'il était illusoire de chercher dans le système visuel un endroit ou se projetterait comme sur un écran de cinéma notre expérience visuelle. Même si V1 occupe une place centrale dans cette architecture entre les entrées thalamiques et les boucles en retour des aires supérieures et que cette aire est essentielle à tout processus visuel supérieur, la représentation globale de l'activité visuelle générée par une image flashée sur la rétine est largement distribuée[†]. Alors que les chemins visuels se mélangent progressivement avec ceux d'autres processus cognitifs, les premières impulsions atteignent le cortex préfrontal puis le cortex moteur après environ $140 - 190$ ms. Sachant que le code est distribué et temporel et connaissant les besoins vitaux correspondant à un réponse rapide aux stimuli visuels, on peut alors se demander quelles sont les contraintes sur ce temps de propagation.

[†]Ainsi, l'IRM montre que l'activité onirique ne nécessite pas forcément une activation des neurones de V1

2.3 Contraintes temporelles sur le code visuel

Nous avons mis en évidence l'importance de la transformation temporelle de l'information dans le système visuel, mais aussi des processus asynchrones qui le régissent. Nous allons dans cette section étudier les contraintes temporelles du code visuel en montrant sa remarquable efficacité. Ces contraintes nous conduiront alors à émettre des hypothèses constructives sur le code visuel.

2.3.1 Catégorisation Ultra-Rapide

Méthodes : présentation sérielle et protocole GO-NOGO

Dans ce but, Thorpe et al. [1996] ont conduit des expériences chez l'homme et le singe rhésus pour déterminer le délai minimum nécessaire pour catégoriser des classes d'images, comme les images contenant ou ne contenant pas un animal. En effet nous avons vu que si on présente rapidement une image, les PAs associés sont propagés le long des chemins visuels pour atteindre le cortex moteur après $\sim 140 - 190$ ms (voir supra) puis enfin par exemple les muscles des doigts après $\sim 180 - 250$ ms. En imaginant le comportement d'une proie qui doit pouvoir reconnaître rapidement le danger d'une scène, on comprend que la pression de l'Évolution sera forte pour garantir les meilleures performances : la contrainte de rapidité est majeure sur le code visuel.
Le protocole de l'expérience s'inspire de protocoles de catégorisation classiques : le sujet, assis devant un moniteur doit porter son regard sur une croix et des images —tirées aléatoirement (sans remise pour éviter un apprentissage) d'une base de donnée d'images naturelles— sont présentées pendant 20 ms, ce qui évite toute saccade pendant la présentation. Dans cette base, la moitié de ces images constitue des cibles (des animaux : poisson, insecte, tigre vu de derrière ou occulté en partie et à n'importe quelle position, posture ou orientation ; voir Fig. 2.3.1-A) et l'autre moitié des distracteurs (comme des paysages, un dirigeable en forme d'éléphant ou des plantes ; voir Fig. 2.3.1-B).

(A) **(B)**

FIG. 2.3.1 – *Catégorisation animal/non-animal : exemples d'images utilisées par [Thorpe et al., 1996]* Les images sont tirées aléatoirement d'une base et représentent ou non un animal : **(A)** Cibles, **(B)** Distracteurs. Les sujets doivent répondre quand une cible est détectée (GO) et seulement dans ce cas (NOGO).

La réponse est donnée selon un processus *GO-NOGO* : les humains doivent relâcher le bouton d'une souris informatique alors que les singes doivent relâcher un bouton et toucher l'écran pour exprimer une réponse positive. Les sujets doivent répondre avec la plus grande précision et le plus rapidement possible à cette tâche de catégorisation. Si l'on peut exprimer explicitement ces conditions expérimentales aux humains, les singes eux sont entraînées par un conditionnement qui associe du jus de fraise à une bonne réponse et une punition (réaffichage de l'image) à une mauvaise (ou une absence) de réponse.

Résultats : la catégorisation a lieu en seulement 150 ms

Cette tâche est effectuée pour n'importe quel type d'animal et requiert donc une classification de haut niveau. Malgré la difficulté de cette tâche (du moins pour un ordinateur), le taux de réussite est élevé chez les humains (94%) mais aussi chez les singes après quelques semaines d'apprentissage (taux de réussite de $87-90\%$). De façon plus surprenante, le temps médian de réponse comprenant la réponse motrice est très rapide : 440 ms chez l'homme et 250 ms chez le singe et se rapproche donc des temps de réponse à des traitement de catégorisation simple (comme indiquer la couleur soit verte soit rouge d'une lampe). On parle alors de *Catégorisation Ultra-Rapide* (CUR).

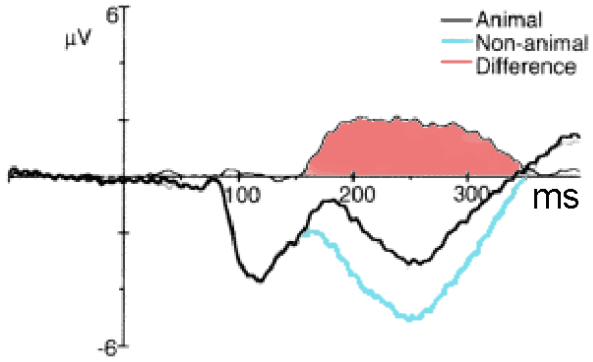

FIG. 2.3.2 – *Activité différentielle entre les potentiels évoqués (EEG) des cibles et des distracteurs.* Trace moyennes pour 15 sujets enregistrées au niveau des électrodes frontales lors de la présentation des cibles (Animal) et des distracteurs (Non-animal). Un divergence très nette se forme dès 150 ms (Difference).

Aussi, l'enregistrement des potentiels évoqués (par l'EEG) permet d'affiner ces résultats. En effet, le potentiel électrique a été enregistré continuellement sur le scalp des sujets pendant ces expériences. S'ils permettent d'évaluer grossièrement la position spatiale de l'activité, l'EEG possède une haute résolution temporelle. En particulier, une divergence très nette apparaît entre les courbes moyennes pour les deux catégories dès 150 ms (voir Fig. 2.3.2). De plus, il a été montré clairement que cette divergence n'était pas un artefact dû à une différence moyenne d'activité, mais qu'elle correspondait bien à la catégorisation des images dans IT [Van Rullen et Thorpe, 2000].

Extensions des résultats

Ces résultats surprenants ont été étendus à de nombreux autres protocoles. Tout d'abord, des tests statistiques ont montré qu'il y avait peu de corrélation entre la latence de la réponse et la configuration de l'image, comme par exemple la taille ou la position de l'animal ou encore son espèce ou la présence de signes distinctifs comme la tête (en particulier les yeux) ou les pattes. Aussi, on peut se demander si la tâche de catégorisation animal/non-animal n'est pas particulière, mais Van Rullen et Thorpe [2000]

a étendu ce protocole à une tâche catégorisant les images contenant des moyens de transports. Même si les singes ne sont naturellement destinés à cette catégorisation, celle-ci a pu être correctement apprise, avec des résultats tout à fait semblables, illustrant de façon étonnante la capacité adaptative du système visuel à des situations nouvelles ou artificielles.
Aussi, de nouvelles expériences neuropsychologiques ont précisé si la manière de présenter l'image n'influençait pas les résultats. Ainsi, Delorme et al. [2000] ont prouvé que les performances ne se dégradaient pas si les images étaient présentées en noir-et-blanc. De plus, il a été aussi montré que ni la familiarité des images [Fabre-Thorpe et al., 2001] ni l'intervalle entre la présentation des images [Delorme, 2001] n'influençaient les résultats.
Enfin, la même expérience a été réalisée dans un environnement différent permettant la projection de l'image en région périphérique. Si l'on observe une dégradation de la réussite, surtout après 70° (les sujets reportent souvent ne pas avoir conscience de l'image mais donnent souvent une réponse significativement plus correcte que le hasard), la rapidité est toujours relativement préservée. De plus, une rapidité similaire est observée pour un protocole où deux images sont présentées en parallèle [Rousselet et al., 2002] et montrent que cette tâche n'est pas traitée séquentiellement mais bien en parallèle. Ces extensions montrent que ce traitement rapide, ce chemin le plus court entre la vision et l'action, suit sûrement la voie Magno (traitement rapide des informations hors de la macula- sans couleur), mais surtout que la CUR révèle un aspect universel et rapide du traitement visuel. Nous allons donc pouvoir formuler des hypothèses quant à la forme de la représentation visuelle.

2.3.2 Conséquences : une première représentation de la scène visuelle par une vague asynchrone de potentiels d'actions

Une propagation *en avant* de l'information visuelle

Ces expériences, appuyées par ce qui a été vu dans Sec. 2.2.2, prouvent que la décision de catégorisation ne peut se faire que dans l'aire inféro-temporale

(IT), la voie tectale limitée au fovéa et insensible à la couleur étant écartée. L'information doit donc suivre depuis la rétine le chemin que nous avons décrit plus haut (voir Fig. 2.2.2) par le CGL, V1, V2, puis la voie temporale ("Quoi ?") jusqu'au cortex inféro-temporal (ce qui explique pourquoi la voie par l'aire MT a été écartée).
Comme nous l'avons vu, la propagation et le traitement de l'information introduisent un délai physique de propagation, et l'on en déduit donc que chaque couche "consommera" une latence d'environ 10 ms par couche [Thorpe et Imbert, 1989]. Ceci ne laisse donc peu de temps dans le passage d'une couche à l'autre et élimine de fait l'influence des chemins *en retour* : l'information du traitement ultra-rapide se propage *en avant* d'aire en aire.

L'information est codée par le premier PA de chaque neurone

Aussi, si on considère que chacune des 10 couches a une même importance dans le processus, alors chaque neurone d'une couche a au maximum 15 ms pour intégrer une réponse pour la couche suivante. Comme la fréquence de décharge des neurones est limitée à environ 200 Hz, on en déduit que les neurones doivent fournir une réponse en intégrant seulement l'information d'au maximum une impulsion par synapse. L'activité sur chaque fibre créée par une image "flashée" doit donc pouvoir être codée dans la première impulsion [Thorpe et Imbert, 1989]. Ce type de codage s'observe de façon explicite dans le système auditif de la chauve-souris [Portfors et Wenstrup, 1999], où chaque neurone répond à la stimulation une fois au plus et avec une grande précision d'autant plus rapidement que l'activité est forte. Mais plus généralement, les enregistrements neurophysiologiques suggèrent que l'activité est codée par la première impulsion et que les impulsions suivantes, et donc par exemple la fréquence de décharge, indiquent une information redondante pouvant potentiellement affiner le codage dans le temps.
Comme on a vu qu'en général un neurone émet d'autant plus rapidement un PA qu'il est activé, la latence de tir d'une impulsion est d'autant plus courte que l'activité est forte : elle peut donc en principe coder une valeur au même titre que la fréquence de décharge qui est l'hypothèse majeure du codage neu-

ronal depuis les découvertes de Adrian [1928]. De plus ce mécanisme privilégie un calcul asynchrone puisque l'activité la plus forte va être mise en avant et transmise en premier. Des données neurophysiologiques appuient cette idée dans le sens où tout d'abord la précision de tir d'un neurone peut être très précise, de l'ordre de la milliseconde [Bair et Koch, 1996], mais aussi très stable [Mainen et Sejnowski, 1996]. Chaque impulsion peut donc potentiellement porter une grande quantité information par sa latence exacte. Aussi, des processus peuvent participer à augmenter la sensibilité des neurones, par exemple en maintenant le potentiel près du seuil de décharge, de telle sorte qu'ils puissent être activés très rapidement.
Aussi, les modulations rythmiques (comme les ondes γ) pourraient jouer un rôle actif dans ce processus. En effet, celles-ci sont corrélées à l'exécution de processus cognitifs comme la catégorisation [Makeig, 1996] et pourraient —si l'on considère ces processus comme le résultat de l'activation d'un système dynamique possédant des fréquences de résonance— modifier en retour de façon rythmique la sensibilité des neurones (par exemple en modifiant le champ local). Ainsi, des décalages de la latence reliés à des mécanismes rythmiques peuvent mettre en avant certaines activités si elles sont corrélées à une activité rythmique [Fries et al., 2001]. On peut alors imaginer qu'à chaque rythme, à l'instant où la sensibilité de groupe des neurones augmente, ce processus amplifie l'expression de la réponse temporelle de l'activité corticale dans une aire en rendant les neurones progressivement plus sensibles, et en les soulignant depuis le plus activé. Ces rythmes illustrerait alors une coopération entre la physiologie et la fonctionnalité du réseau pour pouvoir coder l'image dans l'espace et le temps par un code impulsionnel temporel et asynchrone.

L'image est codée par le front de PAs

Ce modèle est confirmé par des études statistiques poussées sur des enregistrements de neurones *in vivo* initiées par Panzeri et al. [1999]. Ces expériences montrent que si l'on initie une activité transiente, une grande partie de l'information est codée dans les premiers PAs de ce train d'impul-

sions alors que les PAs suivants, organisés en bouffées d'activité ou en un tir régulier, sont plus prédictibles et portent donc par conséquent moins d'information.
L'information visuelle est donc codée —en premier— par le premier PA sur chaque fibre, et donc par le "front" des impulsions neuronales. Ceci a été mis en évidence en particulier dans V1, où les latences de tir du front de PAs sont corrélées de cette façon à la sensibilité à l'orietation [Celebrini et al., 1993]. Ce type de codage est particulièrement robuste à une transmission le long d'une commissure, mais on peut se demander par quels mécanismes celle-ci pourra être décodée.
Ainsi, l'information de synchronie est a priori trop pauvre, puisque quelle que soit sa latence exacte, son information se résumerait à être ou ne pas être synchrone, soit une information d'un bit par neurone qui ne peut pas expliquer le flux d'information observé. En effet, il faudrait alors, par exemple de la rétine au CGL, un nombre bien plus important de fibres. Il semble donc qu'une part supplémentaire de l'information est portée par l'asynchronie des impulsions. Mais comme toutefois il parait improbable qu'il existe une "horloge" globale, les neurones ne peuvent pas "voir" la latence absolue, et nous allons explorer la possibilité que l'information soit codée par la latence relative du front de PAs.

2.3.3 Vers un modèle dynamique de la vision

Nous allons volontairement simplifier ce modèle iconoclaste pour laisser apparaître clairement les processus que ce modèle met en évidence et qui semblent essentiels aux processus visuels et cognitifs.

Un processus en avant et asynchrone...

Tout d'abord à l'opposé des modèles classiques de la vision, ce modèle propose que les processus liés à la CUR, et au traitement visuel rapide en général, sont propagés en avant et de manière asynchrone. C'est à dire tout d'abord qu'un seul passage de l'information par couche est suffisant et que les processus récurrents ne sont pas nécessaires à la réponse. Aussi, le mode

de propagation est événementiel : on suppose alors qu'il n'y a pas d'horloge globale mais que l'information se propage de façon asynchrone et qu'un événement, c'est-à-dire dans ce cas simplifié un simple PA, se propage vers ses efférents en fonction des autres événements et de la configuration du réseau de propagation. Par analogie avec les processus de calcul asynchrone et pour éviter les collisions, on pourra supposer dans ce modèle que le délai de propagation est fixe et que le PA est propagé dès qu'il est émis et définir ainsi un nouveau simulateur de calcul neuronal basé sur des événements [Delorme et Thorpe, 2001b].
Ce mode de fonctionnement a été observé de manière spectaculaire dans le système visuel et les expériences de Keysers et al. [2000] sont à ce titre exemplaire. Ces expériences consistent à présenter à un singe, dont on a trouvé un neurone dans IT sensible à une image "cible" particulière dans une base de donnée, une série d'images de façon très rapide (*Rapid Serial Presentation*) et dans laquelle on introduit (de façon aléatoire) l'image cible. Les auteurs observent alors que même quand la fréquence de présentation des images atteint 71 Hz, soit environ 14 ms par image, le neurone peut répondre à l'image cible. Comme la "distance" en latence de la rétine à IT est d'environ 100 ms (latence qui est confirmé d'autre part par cette expérience), environ 7 vagues d'informations peuvent être traitées en même temps (ici sériellement) comme des fronts successifs d'activation entre la rétine et IT.
Cette expérience suggère donc que les différentes couches transforment de façon très rapide l'information visuelle en des fronts de PAs qui peuvent être traités de façon sérielle dans le système visuel. Ces fronts codent alors l'image sur l'ensemble d'une aire dans un intervalle inférieur à environ 10 ms par l'activation progressive et asynchrone de ses neurones.

... qui étend les modèles existant

Des expériences neurophysiologiques [Nowak et Bullier, 1997], ont montré que la Voie Magno dont nous avons souligné l'importance, pouvait être reliée à des mécanismes de propagation rapide. La vision "utilise" sûrement une voie spécialisée dans l'analyse rapide des scènes [Delorme et al., 1999b] et

ce modèle permettrait donc d'expliquer les processus liés à ce sous-système dans les chemins visuels. Toutefois les fibres Magno se mélangent avec les autres à partir de V1, et donc même si une "première idée" de la scène est envoyée plus rapidement à V1, elle rejoint des chemins communs aux processus visuels après V1.
Tout d'abord, ce modèle de propagation s'intègre dans des architectures dont le fondement est classique (l'organisation du système visuel) et on étendra donc les modèles existants de telle sorte qu'il deviennent compatibles avec la CUR. Par exemple, comme il existe une relation d'équivalence dans un neurone entre la latence et la fréquence de décharge, le code par la latence de première décharge reste compatible avec les modèles classiques, et présente une extension nécessaire pour expliquer la rapidité du système visuel en particulier. Mais il souligne aussi l'importance des processus dynamiques et leurs implication dans les processus cognitifs en général.
Comme on l'a dit, nous allons essayer de simplifier sa formalisation pour en extraire les caractéristiques dynamiques. De possibles généralisations de ce modèle pourront alors étendre le PA à d'autres "quanta" de l'information, comme des signaux plus étendus en temps, comme par exemple des bouffées de PAs [Delorme, 2001], ou alors encore dans l'espace et on peut imaginer que le quanta de l'information soit le tir commun d'une assemblée réduite de neurones (comme ceux d'une colonne corticale).

Un modèle dynamique de la vision

Ce modèle présente une alternative aux codages neuronaux classiques par fréquence, par population ou par synchronie. En effet ceux-ci, à de rares exceptions [van Rossum et al., 2001] ne peuvent prendre en compte la rapidité de la propagation neurale et ne profitent ni de la précision des réponses temporelles ni du caractère asynchrone de la propagation inhérente à un réseau de neurones biologique.
En conclusion de cette étude de la neurophysiologie du système visuel, nous proposons donc un modèle dynamique de la vision où les chemins visuels définissent des voies neuronales de propagation, hiérarchiques et en couches,

depuis la rétine jusqu'à la réponse motrice tel qu'il est formalisé par Thorpe et al. [2001a]. L'information produite par une image brièvement présentée sur la rétine est propagée en avant de manière asynchrone, *i.e.* comme une série d'événements, à travers ce réseau. Le principe majeur de ce type de codage est de montrer qu'en fournisant une réponse impulsionnelle d'autant plus rapidement que l'activité est forte, les neurones rentrent en compétition. Chaque neurone de chaque couche participe ainsi à représenter —temporellement et de façon parallèle— l'information par "ordre d'importance" de telle sorte que les événements les plus saillants soient propagés en avant et que la réponse soit ainsi la plus rapide possible.

Finalement, nous avons défini la vision comme un flux d'information. Nous allons donc présenter des algorithmes qui implantent ces processus pour proposer des modèles de codage neuronal rapide.

Chapitre 3

Modèles de codage rapide

> In a widely circulated joke [from the days of the first computer], a group of engineers assemble the most powerful computer that had ever been conceived and ask it the ultimate question :
> 'Is there a God ?'
> After several tense minutes of clicking and clacking and flashing of lights, a card pops out which reads :
> 'There is *now*.'
>
> *Alwyn Scott* in *How Smart is a Neuron ? A Review of Christof Koch's 'Biophysics of Computation'*

Comme il a été suggéré par Marr [1982] pour construire un modèle de la vision, nous allons maintenant proposer des implantations des définitions neurophysiologiques que nous avons introduit au chapitre 2. Nous allons pour cela introduire tout d'abord des modèles réalistes de neurones à impulsion (voir Gerstner et al. [1999] pour une revue complète) et étudier ainsi les capacités de calcul et d'adaptation des neurones. Nous étudierons alors un modèle de réseau neural impulsionnel simulant la transformation rétinienne et compatible avec les contraintes introduites par les résultats de la catégorisation ultra-rapide (CUR, Sec. 2.3). Enfin, nous allons proposer un modèle de codage neural impulsionnel basé simplement sur l'ordre d'activation des neurones et montrer son efficacité.

3.1 Modèles de codage temporel

3.1.1 Modélisation du mécanisme impulsionnel

Nous allons tout d'abord présenter des modèles de génération, mais aussi d'intégration et d'adaptation des PAs afin d'étudier des modèles impulsionnels de neurones comme base d'un modèle de codage temporel.

Le neurone de Hodgkin-Huxley

Une des plus grandes découvertes scientifiques en neurosciences a été la découverte par Hodgkin et Huxley [1952] du processus cellulaire de génération d'un PA sur l'axone géant du calmar. Celle-ci, récompensée par un prix Nobel, est remarquable autant pour ses résultats que par ses méthodes, et reste aujourd'hui encore une référence. C'est en effet par la confrontation multidisciplinaire sur ce système neuronal simple de mesures neurophysiologiques complètes (extra- et intra-cellulaires et membranaires par "patch-clamp") avec des théories de la physique de l'ingénieur (comme "l'équation du télégraphe" qui régit la propagation d'ondes électriques dans les cables métalliques) qu'ils ont pu modéliser l'évolution du potentiel de membrane par un jeu d'équations différentielles.
Plus précisément, Hodgkin et Huxley ont montré que les courants membranaires pouvaient être décomposés en un courant capacitif (de capacité C) et en différents courants dépendant du potentiel V. En effet, pour chaque canal (ionique : sodium (Na), potassium (K) et de fuite : (L)) le courant est proportionnel à la différence avec le potentiel de repos E du canal et à la conductance du canal $g(V) = \bar{g}.f(V)$, où $f(V)$ est une fonction non linéaire normalisée à 1 pour son potentiel de repos (*i.e.* $g(E) = \bar{g}$), caractérisant la dépendance au potentiel et qu'on note m, n et h.

ion	E (en mV)	$\bar{g}$ (en $^{\mathrm{mS}}/_{\mathrm{cm}^2}$)
Na	115	120
K	-12	36
L	10.6	.3

On obtient alors un système de quatre équations différentielles couplées qui donnent l'évolution du potentiel pour un courant externe d'entrée $\mathcal{I}$.

$$\left\{\begin{array}{rcll} C.\frac{d}{dt}V & = & -\bar{g}_{\mathtt{Na}}.m^3.h.(V-E_{\mathtt{Na}}) & -\bar{g}_{\mathtt{K}}.n^4.(V-E_{\mathtt{K}}) \\ & & -\bar{g}_{\mathtt{L}}(V-E_{\mathtt{L}}) & +\mathcal{I} \\ \frac{d}{dt}x & = & (1-x).\alpha_x(V)-x.\beta_x(V) & \end{array}\right. \tag{3.1.1}$$

où la variable x correspond alternativement aux fonctions auxiliaires m, n et h dont les dynamiques sont couplées à celle de V par les fonctions α_x et β_x. Ces fonctions sont alors pour chaque type de variable des fonctions du potentiel V (en mV), calculées empiriquement par Hodgkin et Huxley pour s'ajuster aux données de l'axone géant du calmar.

x	$\alpha_x(V)$	$\beta_x(V)$
m	$\frac{(2.5-0.1V)}{\exp(2.5-0.1V)-1}$	$4\exp(-u/18)$
n	$\frac{(0.1-0.01V)}{\exp(0.1-0.01V)-1}$	$.125\exp(-u/80)$
h	$0.07\exp(-V/20)$	$\frac{1}{\exp(3-0.1V)+1}$

Les résultats de ce modèle descriptif reproduisent avec une grande précision le comportement de la membrane, la génération du PA et la période réfractaire, ou Post-Hyperpolarisation (voir Fig. 3.1.3-(A)). Ses paramètres sont interprétables comme des mesures statistiques de l'ouverture des canaux ioniques en fonction du voltage, et cette méthode s'étend alors à d'autres types de neurones. D'autres modèles permettent de réduire le nombre de degrés de liberté et rendent le modèle mathématiquement et calculatoirement plus simple. En particulier, le modèle de Morris-Lecar [Rinzel et Ermentrout, 1999] linéarise ces équations et permet un traitement analytique. Si ces modèles permettent la génération intrinsèque d'un PA, ils sont en contrepartie moins généraux et moins précis par rapport au modèle HH. Celui-ci, en expliquant plutôt qu'en copiant une partie des mécanismes neuronaux, est donc le modèle de référence des processus membranaires au niveau de description neuronal.
Par analogie avec la théorie des câbles, il est alors naturel d'étendre ce modèle à des types de neurones plus complexes en considérant qu'il sont composés de *compartiments* réagissant suivant le modèle HH et qu'on "soude" ces différents éléments pour former un neurone complet. Le comportement de ce

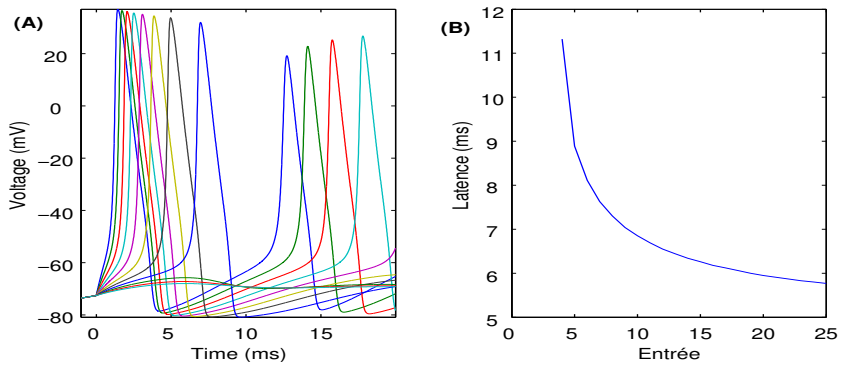

FIG. 3.1.1 – *Modèle impulsionnel de Hodgkin-Huxley.* **(A)** Modèle Hodgkin-Huxley excité à $t = 0\text{ms}$ par des échelons de courant d'amplitudes croissantes. Le neurone répond pour une excitation suffisante et génère alors un PA. Les PAs générés répliquent avec une grande précision le PA observé dans l'axone géant du calmar : la forme d'un PA est relativement indépendant de la stimulation, mais sa latence change avec l'amplitude. Le neurone est ensuite hyperpolarisé mais, comme l'excitation est constante, il peut générer un nouveau PA. **(B)** Latence du premier PA du neurone en fonction de l'amplitude de l'excitation, calculée quand le potentiel dépasse le potentiel -56mV. À partir d'un certain seuil de courant, le neurone génère un PA de plus en plus rapidement selon une fonction monotone.

système peut alors être simulé avec une grande précision mais au prix d'une grande complexité qui rend souvent les calculs de simulation très coûteux. Des logiciels spécialisés comme NEURON [Hines et Carnevale, 1997] ou GENESIS [Protopapas et al., 1999] facilitent toutefois la création de tels modèles. Ils permettent alors de mettre en évidence l'importance du processus d'intégration dendritique ou de mécanismes élémentaires [Sanchez-Vives et al., 2000], comme la différence des constantes de temps de processus inhibiteurs (GABA-A et -B) suggérant leur importance dans les processus neuronaux. En particulier, ils permettent de modéliser par des systèmes d'équations différentielles similaires au modèle Hodgkin-Huxley le comportement d'une synapse grâce au courant qu'elle engendre lors de l'arrivée d'un PA. Ce dernier point permet alors de construire le neurone comme un système complet transformant un train de PAs en entrée en un autre train de PAs de sortie.

Un modèle explicite : l'Intègre-et-Tire à Fuite

Une autre classe de modèles simplifie les équations régissant le neurone en quantifiant plutôt qu'en expliquant des processus comme l'intégration d'un PA par une synapse ou la génération du PA au soma : ces courants sont alors remplacés par une simple règle. Historiquement, le premier modèle, le modèle Intègre-et-Tire (Integrate-and-Fire, IF [Lapicque, 1907]), est aussi un des plus simples : son activité interne est mesurée simplement par son potentiel de membrane V, et à l'image du neurone biologique, comprend une partie intégratrice (l'arbre dendritique), qui va correspondre à une somme linéaire des entrées pondérée par les poids synaptiques et une partie non-linéaire qui simule la génération d'un PA quand V dépasse un seuil : on dit alors que le neurone "tire". Celui-ci est alors inhibé (pendant la période réfractaire) en fixant durant cette période son potentiel à une valeur dépolarisée. Par souci de réalisme, on peut aussi introduire un courant de fuite pour obtenir le modèle LIF (Intégre-et-Tire à Fuite ou *Leaky Integrate-and-Fire*).
Par analogie avec un circuit électrique du premier ordre, le LIF est défini par une équation différentielle du 1° ordre de constante de temps $\tau_m = \frac{C_m}{g_m}$ où g_m et C_m sont resp. la conductance et la capacité membranaire, caractérisant le

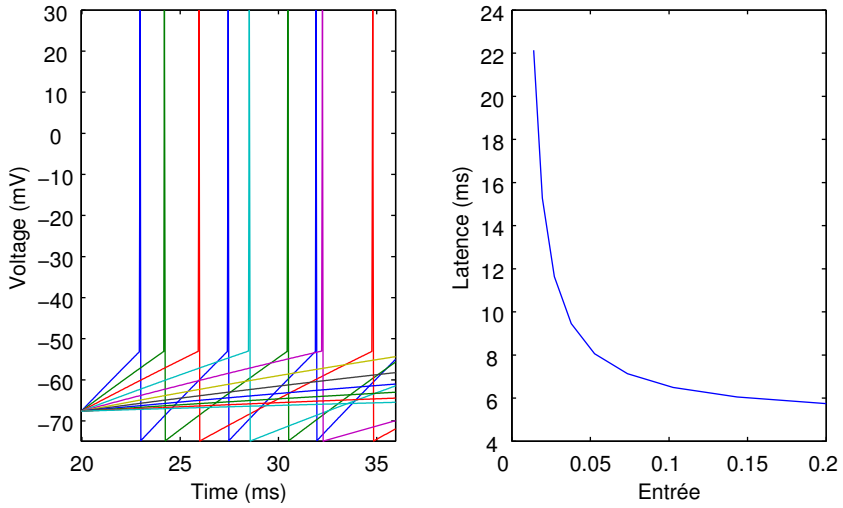

FIG. 3.1.2 – *Modèle de neurone impulsionnel, l'Intègre-et-Tire à Fuite.* **(A)** Le modèle Intègre-et-Tire à Fuite est excité selon le même protocole que pour la Fig. 3.1.1 et en choisissant $\tau_m = 20\text{ms}$, $t_{\texttt{ref}} = 2\text{ms}$, $V_{\texttt{seuil}} = -54\,mV$, $V_{\texttt{repos}} = -70\,mV$ et $V_{\texttt{RAZ}} = -60\,mV$. De façon similaire, le neurone répond pour une excitation suffisante et génère alors un PA et sa latence change avec l'amplitude. De même, le neurone est ensuite hyperpolarisé et peut générer un nouveau PA. **(B)** Latence du premier PA du neurone en fonction de l'amplitude de l'excitation. À partir d'un certain seuil de courant, le neurone génère un PA de plus en plus rapidement selon une fonction monotone similaire au modèle de Hodgkin-Huxley.

---

temps de fuite. Le potentiel de membrane $V(t)$ dévie alors du potentiel de repos $V_{\texttt{repos}}$ en intégrant le courant d'entrée $\mathcal{I}(t)$ jusqu'à atteindre le seuil. La sortie est alors la suite des PAs $\{t_i^k\}$, $k \geq 0$.

$$\left\{ \begin{array}{l} C_m.\frac{dV_i}{dt}(t) = -g_m.(V_i(t) - V_{\texttt{repos}}) + \mathcal{I}_i(t) \\ \text{si } V_i(t) > V_{\texttt{seuil}}, \text{ alors } \left\{ \begin{array}{l} t_i^{(f)} := t \\ \text{ et } V_i(t_i^k < t < t_i^k + t_{\texttt{ref}}) = V_{\texttt{RAZ}} \end{array} \right. \end{array} \right. \tag{3.1.2}$$

Les potentiels $V_{\texttt{seuil}}$ et $V_{\texttt{RAZ}}$ correspondent resp. aux potentiels de seuil et de remise à zéro. Par conséquent, si au temps $t = 0$ on applique un courant constant d'amplitude $\mathcal{I}_0$ au neurone à son potentiel de repos ($V = V_{\texttt{repos}}$), on peut calculer explicitement le comportement linéaire du neurone pour les

temps positifs.

$$V(t) = \frac{\mathcal{I}_0}{C_m}.[1 - \exp(-\frac{t}{\tau_m})] \tag{3.1.3}$$

Par conséquent, le neurone répondra seulement pour $\mathcal{I}_0 > (V_{\texttt{seuil}} - V_{\texttt{repos}}).g_m$ par un premier PA à la latence

$$t(\mathcal{I}_0) = \tau_m.\ln(\frac{\mathcal{I}_0}{\mathcal{I}_0 - (V_{\texttt{seuil}} - V_{\texttt{repos}}).g_m}) \tag{3.1.4}$$

puis régulièrement à la fréquence $f(\mathcal{I}_0) = \frac{1}{t(\mathcal{I}_0)}$. De la même manière que la modèle HH (voir Fig. 3.1.1-(B)), ce modèle fournit donc une transformation simple d'une valeur analogique en une latence et en une fréquence de décharge. En particulier, on observe que cette valeur est codée de façon bijective par la latence du premier PA, par l'intervalle inter-impulsionnel et par la fréquence de décharge. Ce modèle doit maintenant être étendu pour pouvoir modéliser le rôle des synapses.

Modèles de réponse synaptique

De la même manière que nous avons défini la transformation d'un courant en un signal impulsionnel par analogie avec un circuit électrique du 1° ordre, traduisons l'effet d'un PA sur une synapse par un courant dans le neurone post-synaptique (voir [Destexhe, 1997] pour une revue) afin de construire un modèle complet de neurone impulsionnel. Notons S_j l'entrée provenant d'un neurone émetteur j sur un neurone récepteur i et qui est transmise à travers la synapse (ij). Tout d'abord, nous modéliserons le signal impulsionnel, le *tarin de spikes*, entrant sur la synapse (ij) comme un somme de distributions de Dirac*. On notera Γ_j la liste de PAs correspondant à des temps de tir t_j^k tirés du neurone émetteur.

$$S_j = \sum\nolimits_{t_j^k \in \Gamma_j} \delta(t - t_j^k) \tag{3.1.5}$$

*Un Dirac est un modèle mathématique représentant une fonction de somme des aleurs absolues (norme L_1) normalisée à 1 mais dont l'étalement temporel est négligeable : sa localisation temporelle est alors considérée comme ponctuelle.

On considérera deux catégories de synapses, excitatrices et inhibitrices (sous-ensembles de l'ensemble des synapses que nous noterons simplement `exc` et `inh`), que nous caractériserons ici par le courant synaptique qu'elles génèrent, mais agissant sur le potentiel du neurone par des canaux ioniques et des neurotransmetteurs différents, et donc avec des potentiels différents (resp. $E_{\texttt{exc}}$ et $E_{\texttt{inh}}$). En négligeant dans ce modèle l'effet de l'arbre dendritique, on approche alors avec une grande précision le courant synaptique excitant le neurone comme la réponse de systèmes dynamiques linéaires du premier ordre de constante de temps $\tau_{\texttt{syn}} \sim 5$ms aux entrées synaptiques S_j. Chaque synapse est alors caractérisée par son type et son poids w_{ij} (de dimension physique un courant) et le courant synaptique à la synapse (ij) est déterminé par :

$$\frac{d}{dt}g_{ij}(t) = -\frac{1}{\tau_{\texttt{syn}}}.g_{ij}(t) + w_{ij}.S_j \tag{3.1.6}$$

et le courant total par

$$\mathcal{I}_i(t) = g_i^{\texttt{exc}}.(V - E_{\texttt{exc}}) + g_i^{\texttt{inh}}.(V - E_{\texttt{inh}}) \tag{3.1.7}$$

$$\text{avec} \quad \left\{ \begin{array}{rcl} g_i^{\texttt{exc}} & = & \sum_{(ij)\in\texttt{exc}} g_{ij} \\ g_i^{\texttt{inh}} & = & \sum_{(ij)\in\texttt{inh}} g_{ij} \end{array} \right. \tag{3.1.8}$$

ce qui définit entièrement un modèle de neurone LIF à courants synaptiques (LIF-SC) comme une transformée d'un train de PAs en un train de PAs. En résumé, de Éq. 3.1.2 et Éq. 3.1.8, il vient pour le comportement linéaire du modèle :

$$\frac{d}{dt}V_i = -\frac{1}{\tau_m}.(V_i - V_{\texttt{repos}}) + g_i^{\texttt{exc}}.(V_i - E_{\texttt{exc}}) + g_i^{\texttt{inh}}.(V_i - E_{\texttt{inh}})$$

On peut alors définir l'effet d'un PA reçu à $t = t_0$ pour une synapse avec un poids $w = 1$ unitaire en intégrant le courant dans Éq. 3.1.6 avec $S_j = \delta(t - t_0)$. Ce calcul permet de calculer directement le Courant Post-Synaptique :

$$g(t) = [1 - exp(-\frac{t - t_0}{\tau_{\texttt{syn}}})].\Delta(t \geq t_0) \tag{3.1.9}$$

où $\Delta(t)$ est la fonction échelon ($\Delta(t) = 0$ pour $t < 0$ et 1 sinon). De la même manière, si l'on néglige pour le courant $\mathcal{I}_i(t)$ dans Éq. 3.1.8 de la variation de potentiel, il vient pour le potentiel synaptique que nous dénoterons ε dans la suite

$$\varepsilon(t) = \frac{\tau_m}{\tau_m - \tau_{\mathtt{syn}}}.[exp(-\frac{t-t_0}{\tau_m}) - exp(-\frac{t-t_0}{\tau_{\mathtt{syn}}})].\Delta(t \geq t_0) \quad (3.1.10)$$

La pente de ε vaut donc 1 pour $t = t_0$. Ce modèle définit bien le Potentiel Post-Synaptique (voir Sec. 2.1.2) comme la réponse normalisée d'un PA sur le potentiel de la membrane (voir Fig. 3.1.3-(B) avec $t_0 = 0$). La confrontation avec des mesures neurophysiologiques permettent de définir $\tau_{\mathtt{syn}} \sim 5$ms. Cette fonction atteint son maximum pour $t = \frac{\tau_m.\tau_{\mathtt{syn}}}{\tau_m - \tau_{\mathtt{syn}}}.\ln(\frac{\tau_m}{\tau_{\mathtt{syn}}})$, soit environ $t = 9.24$ ms. On pourra définir la force d'une synapse par la pente du PSP à son origine, c'est-à-dire que le poids synaptique pourra se déduire d'un enregistrement neurophysiologique par la pente du potentiel post-synaptique. Ainsi, on pourra définir simplement le PSP en multipliant ε par le poids synaptique.

Neurone à réponse impulsionnelle

Une extension de ce modèle est le modèle du *neurone à réponse impulsionnelle* (Spike Response Model ou SRM, [Gerstner et al., 1999, p.16]). Le principe de ce modèle est de mémoriser la réponse impulsionnelle, c'est-à-dire la réponse à un PA de la synapse (le PPS) depuis un modèle (e.g. Hodgkin-Huxley ou LIF-CS, en négligeant la dépendance au voltage, voir Sec. 3.1.1) ou depuis des enregistrements neurophysiologiques. Comme dans le LIF, on considère que la dendrite intègre ces contributions en les sommant, *i.e.* qu'on peut calculer son potentiel comme la somme modulée par les poids synaptiques. De façon similaire, quand le potentiel atteint le potentiel de seuil, le neurone génère un PA et on ajoute au potentiel un potentiel tabulé qui correspond au PA et à période réfractaire comme un potentiel η. Par analogie avec le modèle LIF, on peut par exemple choisir η tel que si le neurone i tire

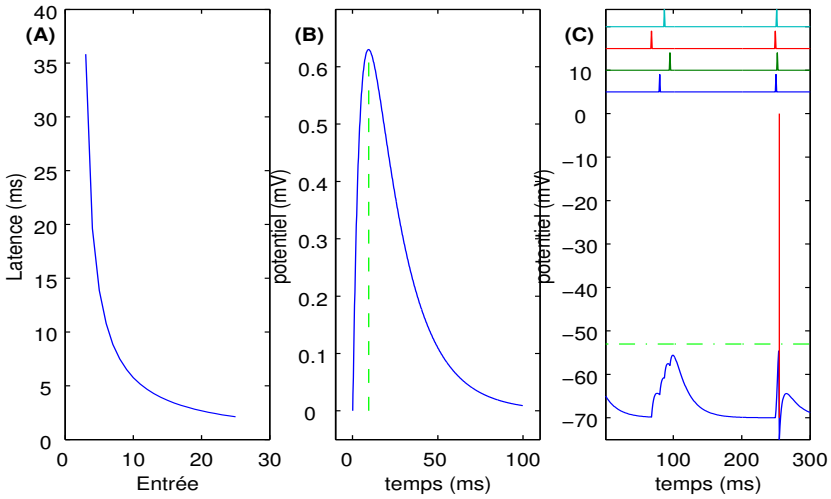

FIG. 3.1.3 – *Modèle intègre-et-tir à réponse impulsionnelle.* **(A)** Latence du premier PA pour le modèle SRM en fonction de l'amplitude de l'entrée avec le même protocole que pour la Fig. 3.1.1. À partir d'un certain seuil de courant, le neurone génère un PA de plus en plus rapidement, à l'image du comportement du modèle HH (voir Fig. 3.1.1-(B)). **(B)** Modèle d'impulsion synaptique : le Potentiel Post-Synaptique. Les constantes de temps du modèle sont fixées à 5ms pour la synapse et 20ms pour la membrane. Un PA arrive sur la synapse à $t = 0$, provoquant un potentiel dont la forme est toujours la même (maximum à $t = 9.2\text{ms}$ mais dont l'amplitude varie proportionnellement avec l'efficacité synaptique. **(C)** Modèle LIF-SC : intégration de PAs (présentés en haut du graphe) sur 4 synapses équipotentes. L'entrée est constituée de PAs sur chaque synapse avec une synchronie faible à $t = 100\text{ms}$ et forte à $t = 250\text{ms}$. La réponse favorise les entrées synchrones.

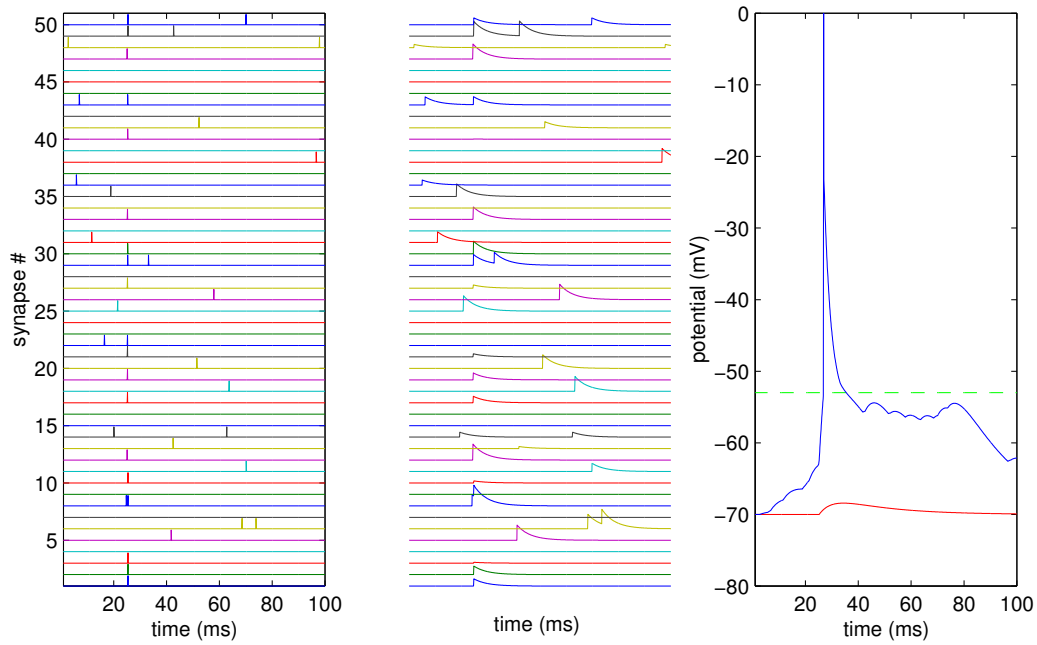

FIG. 3.1.4 – *Modèle de neurone impulsionnel.* On a modélisé un neurone SRM avec 50 entrées synaptiques. **(Gauche)** Trains de PAs en entrée, générés par un processus de Poisson et auquel on a ajouté à $t = 25\text{ms}$ un motif synchrone sur la moitié des synapses (avec un étalement temporel de 2ms). **(Milieu)** Courants synaptiques dans le temps. **(Droite)** Potentiel du neurone, un PPS (généré par la synapse 1) est montré comme illustration. Le neurone émet un PA quand il dépasse le seuil (ligne discontinue).

à $t = 0$ (voir Éq. 3.1.2) :

$$\begin{cases} \tau_m . \frac{d}{dt}\eta(t) = -\eta(t) \\ \text{avec } \eta(0) = V_{\texttt{RAZ}} - V_{\texttt{seuil}} \end{cases} \tag{3.1.11}$$

soit

$$\eta(t) = (V_{\texttt{RAZ}} - V_{\texttt{repos}}).exp(-\frac{t}{\tau_{\texttt{m}}}).\Delta(t \geq 0) \tag{3.1.12}$$

Étant donnés ε et η, les potentiels correspondant resp. au PA et à la période

réfractaire, le modèle SRM est alors donné par :

$$\begin{cases} V_i(t) = \eta_i(t - t_i^{(\mathtt{f})}) + \sum_j \sum_{t_j^k \in \Gamma_j} w_{ij}\varepsilon(t - t_j^k) \\ \text{si } V_i(t) > V_{\mathtt{seuil}} \text{ et } V_i(t - \delta t) < V_{\mathtt{seuil}}, \text{ alors } t_i^{(\mathtt{f})} := t \end{cases} \tag{3.1.13}$$

Bien que seulement descriptif, ce modèle peut simuler avec précision la réponses d'un neurone biologique. De plus, ce modèle est linéaire et, contrairement au modèle HH, permet une implantation et une étude mathématique simple de la transformation du neurone. Il permet aussi par rapport au modèle LIF de décrire sans équations différentielles un comportement similaire en réduisant ainsi le nombre de variables en les remplaçant par des variables exogènes. En fait, ces deux modèles sont exactement équivalents si l'on choisit ϵ et ν correctement comme les réponses impulsionnelles de la synapse et de l'axone lors de la génération d'un PA et si l'on s'abstrait de la dépendance des synapses au potentiel de membrane, comme défini au paragraphe précédent.
Un grand avantage de ces deux derniers modèles par rapport au modèle HH, en réduisant les calculs mais surtout en construisant des règles événementielles (réception ou émission d'un PA). Ces modèles sont facilement implantables sur des architectures de type VLSI [Mead, 1989] qui permettent d'envisager des applications en temps réel. Il permettent aussi d'étudier les conséquences de l'inclusion de mécanismes plus précis ou encore de créer des modèles de réseaux de neurones. Dans le reste de cet ouvrage, nous utiliserons le modèle LIF-SC avec les constantes que nous avons données dans Fig. 3.1.2 mais nous utiliserons aussi une formalisation équivalente par SRM pour étudier le comportement mathématique de ces neurones.

Processus non-linéaires élémentaires du code neuronal

> Thinking about brain style computation requires a certain frame of mind, related to but distinctly different from that of the biophysicist. For instance, how should we think of a chemical synapse ? In terms of complicated pre- and post-synaptic elements ? Ionic channels ? Calcium binding proteins ? Or as a non-reciprocal and stochastic switching device that transmits a binary signal rapidly between two neurons and remembers its history of usage ? The

> answer is that we must be concerned with both aspects, with biophysics as well as computation.
>
> *Christof Koch* [Koch, 1998]

En général, les neurones ont une morphologie et des fonctions plus complexes que celles de l'axone géant du calmar, et au lieu d'investir une grande quantité de calcul dans la précision d'un modèle étendu comme le modèle HH, il est souvent plus important de prendre en compte les nombreux processus non-linéaires qui interviennent dans le traitement, comme autant de "briques" de l'algorithme neural.

Dans l'arbre dendritique par exemple, on peut introduire directement le délai correspondant à la propagation de la synapse au soma, ou encore considérer que l'intégration des entrées est non-linéaire[†]. On peut aussi intégrer des non-linéarités "fortes" comme celle introduite par une règle comme le seuillage dans le modèle IF. Koch et Segev [2000] a dans ce sens proposé un modèle où si deux "branches" de l'arbre dendritiques qui se rejoignent sont activées en même temps, le résultat propagé est supérieur à la somme de leurs contributions, implantant ainsi dans l'arbre dendritique le calcul de fonctions temporelles non linéaires. Ce comportement, qui peut être rencontré dans des neurones pyramidaux du cortex permet alors d'effectuer des calculs algébriques complexes.

D'autres processus non-linéaires ont été observés au niveau des boutons synaptiques. En particulier, les synapses chimiques semblent jouer un rôle clé dans ces processus et apparaissent alors comme des systèmes dynamiques complets d'une complexité équivalente au neurone [Tsodyks et al., 1998]. Comme nous l'avions évoqué, le signal synaptique est transmis par l'intermédiaire de neurotransmetteurs, contenus dans des vésicules (voir Fig. 3.1.5) qui s'accumulent dans le site pré-synaptique, mais qui peuvent être en nombre limité. Leur libération est stochastique (de l'ordre de moins de 30% de chance d'être libéré lors de l'arrivée d'un PA). On peut alors développer un modèle de la dynamique synaptique à partir de l'estimation de la probabilité $p_{\texttt{pre}}$ d'émission d'une vésicule en fonction de l'entrée pré-synaptique (les PAs)

[†]La théorie des câbles cylindriques prévoit ainsi que la linéarité est en $\sqrt[3]{V^2}$

[Senn et al., 1997]. De la même façon, l'activation des récepteurs post-synaptiques à laquelle correspond la probabilité $p_{\texttt{post}}$ peut être modélisée en fonction des PAs se rétro-propageant jusqu'à la synapse.
Ce modèle étend le modèle LIF-CS que nous avons présenté plus haut en tenant compte de l'influence du nombre limité de ressources et donc de l'historique d'activation aux niveaux pré- et post-synaptiques. Ce modèle présente la réponse d'une synapse à un PA isolé et ne tient pas compte de tout l'historique de tir de i ou j. L'efficacité synaptique $\bar{g}$ va ici dépendre du produit d'une efficacité au repos $\bar{g}_0$ modulé par des facteurs d'adaptation de transmission pré- et post-synaptiques $A_{\texttt{pre}}$ et $A_{\texttt{post}}$ qui vont être modifiés par l'activité temporelle de la synapse —et qui donc au repos valent 1. Formellement, on peut écrire

$$\bar{g} = \bar{g}_0.A_{\texttt{pre}}.A_{\texttt{post}} \tag{3.1.14}$$

On déduira l'évolution dynamique de l'efficacité $\bar{g}$ par les équations différentielles de la cinétique d'un réaction chimique du premier ordre correspondant à des densités de populations de processus (vésicules pré-synaptiques et canaux post-synaptiques libres) limités en nombre et réagissant à un médiateur.

$$\text{(Libre et Non-Actif)} \; + \; \text{Médiateur} \rightleftharpoons \text{(Occupé et Actif)} \tag{3.1.15}$$

Cette réaction est initiée par les PAs (qui modifient alors l'équilibre Éq. 3.1.15 de gauche à droite) et on peut donc écrire les équations qui dirigent les variations des facteurs A suivant les temps d'arrivée de PAs pré-synaptiques pour $A_{\texttt{pre}}$ ou de PAs post-synaptiques rétrogrades pour $A_{\texttt{post}}$ et suivant leur constantes de temps de récupération respectives $\tau_{\texttt{pre}}$ et $\tau_{\texttt{post}}$ (de droite à gauche dans l'équilibre Éq. 3.1.15).

$$\begin{cases} \frac{dA_{\texttt{pre}}}{dt} = \frac{1}{\tau_{\texttt{pre}}}.(1 - A_{\texttt{pre}}) + \alpha_{\texttt{pre}}(\sum_j \delta(t - t_j)).(A_{\texttt{pre}}^{\texttt{actif}} - A_{\texttt{pre}}) \\ \frac{dA_{\texttt{post}}}{dt} = \frac{1}{\tau_{\texttt{post}}}.(1 - A_{\texttt{post}}) + \alpha_{\texttt{post}}(\sum_i \delta(t - t_i)).(A_{\texttt{post}}^{\texttt{actif}} - A_{\texttt{post}}) \end{cases} \tag{3.1.16}$$

Les paramètres α (généralement de l'ordre de $0,3$) décrivant l'effet relatif (ils sont donc compris entre 0 et 1) d'un PA, ce qui correspond aux observations

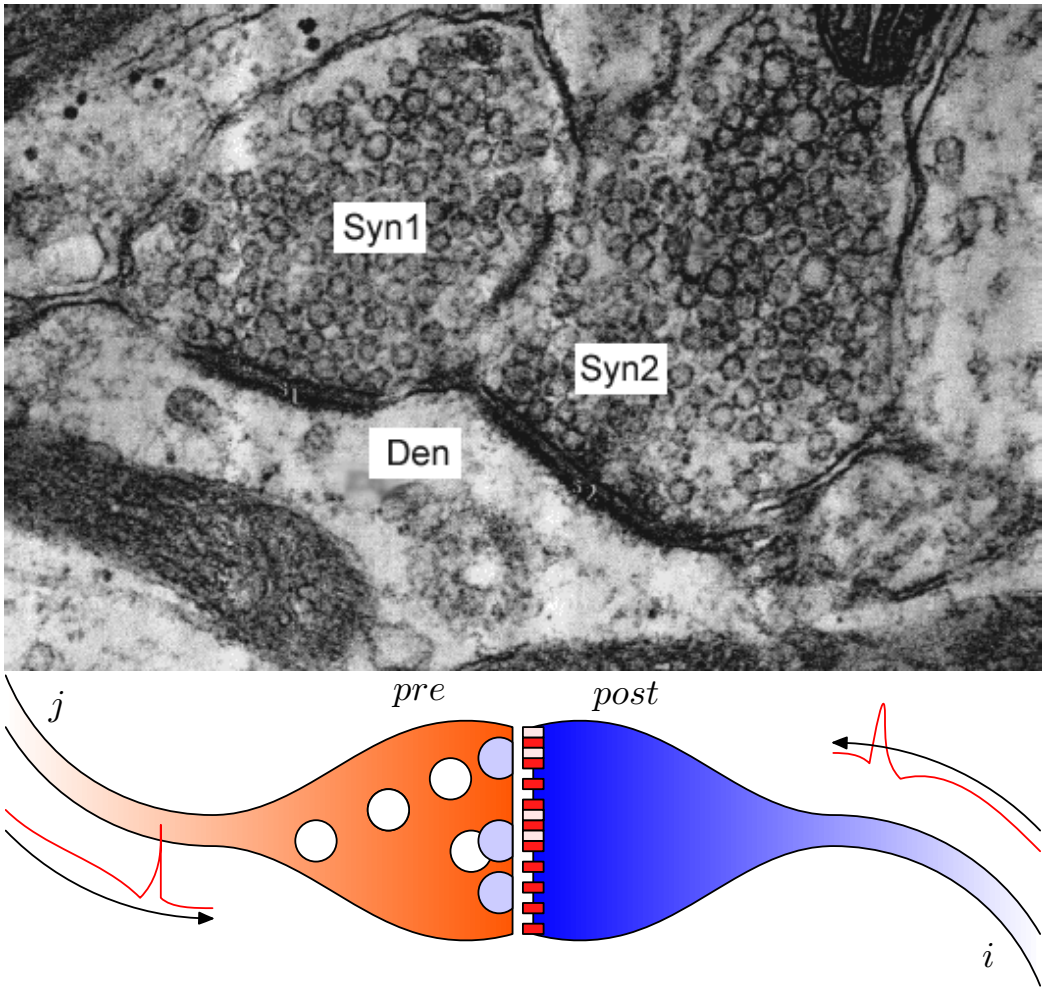

FIG. 3.1.5 – *Synapses.* **(A)** Cette microphotographie présente deux synapses (Syn1 et Syn2) sur une dendrite (Den) du cortex auditif du rat. Elle met en évidence les vésicules de neurotransmetteur ainsi que l'espace inter-synaptique plus sombre. (source : [Burt, 1992, p.42]) **(B)** Schéma du modèle d'une synapse entre un neurone émetteur (pre) j et un neurone récepteur (post) i. Les disques correspondent aux vésicules pré-synaptiques et les couleurs représentent leur disponibilité (facteur $A_{\texttt{pre}}$). Les rectangles qui leur font face correspondent alors aux canaux sensibles à ce neurotransmetteur alors que leur couleur correspond à la densité de canaux recrutés pour la transmission (facteur $A_{\texttt{post}}$). En écrivant l'équilibre de ces processus suivant leurs activations par les PAs, on pourra déterminer les variations de l'efficacité de la synapse et donc l'amplitude du potentiel post-synaptique qui va exciter le neurone.

neurophysiologiques. Les valeurs $A^{\texttt{actif}}$ décrivent alors la valeur asymptotique quand dans Éq. 3.1.15 tous les processus sont actifs. Au premier ordre une facilitation forte ($A^{\texttt{actif}} >> 1$) correspondra à une addition de l'efficacité à chaque PA, alors qu'une dépression ($0 < A^{\texttt{actif}} << 1$) correspondra à une division de l'efficacité.
Même si ce modèle est très simplifié par rapport à des modèles tenant compte des mécanismes synaptiques (comme le comportement d'ancrage des vésicules sur la membrane pré-synaptique), les simulations correspondant à ce modèle sur un neurone LIF-CS montrent qu'à court terme ces processus élémentaires vont avoir une forte influence sur la forme du train de PAs. En particulier, ils permettent d'interpréter physiologiquement des phénomènes de *facilitation* et de *dépression* qui sont observés dans les expériences neurophysiologiques. En particulier, le processus rétrograde, en se propageant à l'ensemble des synapses d'un neurone, joue un rôle de contrôle du gain [Abbott et al., 1997] ou de processeur temporel [Chance et al., 1998]. Aussi, Markram et Tsodyks [1996] ont prouvé que ce mécanisme pouvait dans les neurones être relié à une redistribution des poids synaptiques, et donc à un mécanisme d'adaptation à court terme. Ce dernier mécanisme montre qu'alors de telles "briques élémentaires" du processus neuronal peuvent par exemple amplifier le contraste de la composante spatiale plutôt que temporelle d'un signal convergeant vers le soma, et qu'elles impliquent donc une grande variation qualitative du résultat (voir Fig. 3.1.6).

3.1.2 Codage impulsionnel et adaptation synaptique

L'importance des mécanismes synaptiques montrent que le neurone n'est donc pas la seule unité computationnelle dans le SNC. Nous allons préciser le rôle de la synapse en étudiant son adaptation en fonction du flux impulsionnel qui la traverse.

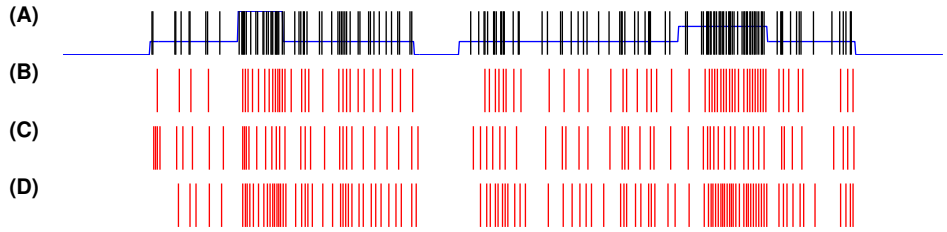

FIG. 3.1.6 – *Facilitation et dépression.* Pour étudier les mécanismes synaptiques à court terme, on a généré un **(A)** train de PAs (traits noirs) par un processus ponctuel dont l'intensité est constante par morceaux (ou "Poisson par morceaux", la fréquence du processus varie de 0 à 20 Hz et est indiquée par la ligne continue) servant d'entrée commune à différents modèles. **(B)** Un modèle LIF avec courants synaptique génère un train de PAs similaire à l'entrée avec un délai constant. **(C)** et **(D)** Train de PAs généré par un modèle LIF avec des mécanismes synaptiques de facilitation et de dépression ($\tau_{\texttt{pre}} = 25\text{ms}$, $tau_{\texttt{post}} = 100\text{ms}$, $\alpha_{\texttt{pre}} = 0,15$ et $\alpha_{\texttt{post}} = 0,5$). Les trains font alors respectivement apparaître des bouffées de PAs très rapides et contrastées avec une facilitation forte (pour $A^{\texttt{actif}}_{\texttt{pre}} = 4$ et $A^{\texttt{actif}}_{\texttt{post}} = 0$) ou une activité plus régulière avec la dépression ($A^{\texttt{actif}}_{\texttt{pre}} = 0,5$ et $A^{\texttt{actif}}_{\texttt{post}} = 2$).

Plasticité hebbienne à modulation temporelle relative (Spike-Time Dependant Plasticity, STDP)

Alors que l'on considère souvent que l'adaptation synaptique résulte d'activations impulsionnelle massives, Debanne et al. [1995] ont prouvé que si on excite une synapse par des PAs pré- et post-synaptiques dont le délai relatif (de l'ordre de 10 ms) est constant et que l'on répète cette expérience, le temps relatif entre ces PAs peut changer à long terme (de l'ordre de 10 min) l'efficacité de cette synapse. En particulier, Markram et al. [1997] et [Bi et Poo, 1998] ont montré que le signe de cette variation dépendait de l'ordre de tir (entre l'impulsion pré et post, voir Fig. 3.1.7) et que l'amplitude de ce changement augmentait avec la brièveté de ce délai (voir Fig. 3.1.8-(A)). Ainsi, si le neurone pré-synaptique contribue à faire tirer le neurone post-synaptique, la synapse est augmentée (Potentiation à Long Terme), alors que si le neurone tirait avant qu'il reçoive un PA, la synapse est diminuée (Dépression à Long Terme). Ce processus va donc privilégier les synapses qui entretiennent une relation causale entre l'entrée et la sortie. On peut donc rapprocher cette règle de la règle de Hebb (voir Sec. 2.1.2), puisque la formulation est qualitativement exacte mais qu'elle s'étend ici à la temporalité des événements.

Abbott et Nelson [2000] ont montré qu'en fait, cette règle d'apprentissage pouvait prendre des formes variées dans des préparations différentes, et particulièrement dans la "fenêtre" correspondant au changement de l'efficacité en fonction du délai. Les conséquences computationnelles du choix de cette fenêtre sont très fortes sur la répartition finale des poids synaptiques et aussi sur la "forme" du motif de décharge des neurones [Song et al., 2000]. La forme de la fenêtre peut donc correspondre à des fonctions neuronales différentes, la configuration des neurones étant une conséquence de ce paramètre. Par exemple, Feldman [2000] montre qu'une fenêtre plus large que Fig. 3.1.8-(A) assure au neurone une adaptation à une privation d'entrées synaptiques (lors de la perte d'un œil, par exemple).

Aussi, Bi et Poo [1998] a par exemple montré que dans certains neurones, si la potentiation est multiplicative, la dépression est soustractive (voir Fig. 3.1.8-(B)), une modification "algorithmique" qui modifie énormément la stabilité

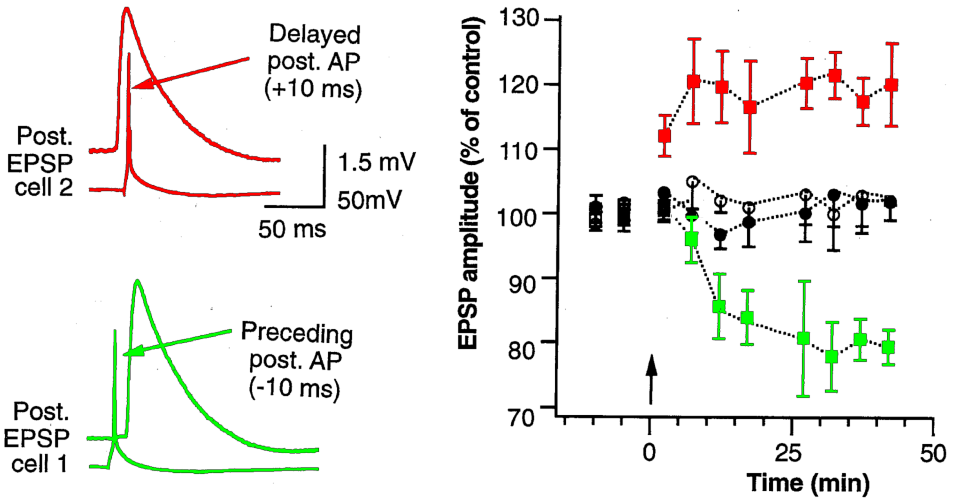

FIG. 3.1.7 – *Plasticité à modulation temporelle relative : expérience.* **(Gauche)** Une synapse entre deux neurones est excitée de façon répétée dans le temps par un PA pré-synaptique (qui génère un PPS, trace du haut) et par un PA post-synaptique. On provoque à chaque présentation l'émission de ces PAs avec un délai fixe (ici du PA post-synaptique par rapport au PPS) qui est retardé (Haut) ou avancé (Bas). **(Droite)** Modulation de l'efficacité de la synapse (mesurée par la pente du PPS) en fonction du délai et dans le temps. Après environ 10 minutes, on observe que la synapse correspondant à l'association causale (soit la succession PA pre puis PA post) est renforcée alors que l'efficacité de l'autre synapse est diminuée (adapté de Markram et al. [1997]).

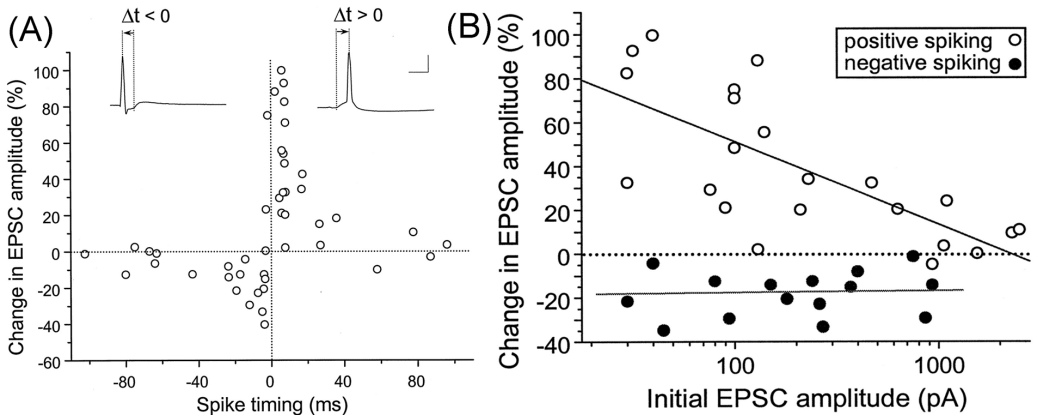

FIG. 3.1.8 – *Plasticité à modulation temporelle relative : résultats.* **(A)** Variation de l'efficacité de la synapse en fonction du délai. Si l'on répète l'expérience pour des délais différents, le signe de la variation reste similaire à Fig. 3.1.7 : les synapses causales sont renforcées. Mais l'amplitude de la variation diminue avec le délai : des délais plus faibles correspondent à un apprentissage plus rapide. **(B)** Variation de l'efficacité relative en fonction de l'amplitude de l'efficacité. Si l'on trace maintenant la variation en fonction de l'efficacité originale, on observe que la potentiation est multiplicative et la dépression est soustractive. (source : Bi et Poo [1998])

[Xie et Seung, 2000] et la distribution finale des poids [Kempter et al., 1999]. Aussi, des modèles explicites (avec une forme de fenêtre fixée) ont été étudiés intensivement pour en déduire l'évolution de ses poids synaptiques et l'influence sur la stabilité de l'apprentissage [Rubin et al., 2000; Rossum et al., 2000]. En particulier, Cateau et Fukai [2002] a montré de façon générale l'importance de la forme de cette fenêtre temporelle et propose un algorithme "inverse" déduisant cette fenêtre de la fonction recherchée.

Modèle génératif de plasticité

On peut relier ces mécanismes d'adaptation à long terme aux mécanismes synaptiques que nous avons évoqué plus haut pour en déduire un modèle de la plasticité. De la même manière que nous avions modélisé le potentiel et les courants par des cinétiques du premier ordre, on peut en effet modéliser la plasticité de la synapse par sa constante de temps τ_g. Celle-ci est plus

longue que pour les règles précédentes puisqu'elle peut être de l'ordre de la seconde ou de la minute. On en déduira alors un modèle dynamique de la synapse contrôlé par l'historique d'arrivée des PAs sur celle-ci. Par analogie avec les mécanismes de facilitation et de dépression, on pourra relier cette modification aux probabilités évaluées précédemment et écrire la variation de l'efficacité synaptique brute $\bar{g}_0$ entre 0 et une valeur maximale $\bar{g}_{\mathtt{max}}$ correspondant biologiquement à la densité maximale de médiateurs synaptiques (vésicule et canaux). Notons alors $\bar{g}_0 = \bar{g}_r.\bar{g}_{\mathtt{max}}$, et donc $\bar{g}_r \in [0,1]$ est l'efficacité brute relative à son maximum. Alors, comme décrit dans [Perrinet et Samuelides, 2000], on peut associer un changement synaptique à chaque arrivée de PA sur la synapse. Son efficacité varie relativement à sa valeur vers 1 (renforcement) quand un PA rétrograde arrive sur la synapse et vers 0 (affaiblissement) quand un PA arrive d'un neurone émetteur.

$$\begin{aligned} \tau_w \frac{d\bar{g}_r}{dt} = \quad & +(1-\bar{g}_r). \quad (A_{\mathtt{pre}}-1). \sum_l \delta(t-t_l^{\mathtt{post}}) \\ & -\bar{g}_r.(A_{\mathtt{post}}-1). \sum_k \delta(t-t_k^{\mathtt{pre}}) \end{aligned} \tag{3.1.17}$$

En particulier, si on réplique le protocole d'excitation pré- et post-synaptique de l'expérience de Markram et al. [1997], alors dans le cas d'un tir pré-synaptique à $t = 0$ suivi d'un tir post-synaptique à $t_{\mathtt{post}}$, il vient $A_{\mathtt{pre}} = 1 + K.\exp(-\frac{t}{\tau_{\mathtt{pre}}})$ avec $K = \alpha_{\mathtt{pre}}.(A_{\mathtt{pre}}^{\mathtt{actif}} - 1)$ et donc

$$\Delta\bar{g}_r \quad = \quad +(1-\bar{g}).K.\exp(1-\frac{t_{\mathtt{post}}}{\tau_{\mathtt{pre}}})$$

De la même façon pour un tir post-synaptique à $t = 0$ suivi d'un tir pré-synaptique à $t_{\mathtt{pre}}$, il vient pour $K' = \alpha_{\mathtt{post}}.(A_{\mathtt{post}}^{\mathtt{actif}} - 1)$

$$\Delta\bar{g}_r \quad = \quad -\bar{g}.K'.\exp(1-\frac{t_{\mathtt{pre}}}{\tau_{\mathtt{post}}})$$

Nous avons donc fait ici une approximation de la fenêtre d'apprentissage par des exponentielles décroissantes. On peut alors reproduire les résultats de la Fig. 3.1.8-(A) avec les paramètres adéquats (voir Fig. 3.1.9-(A)). On

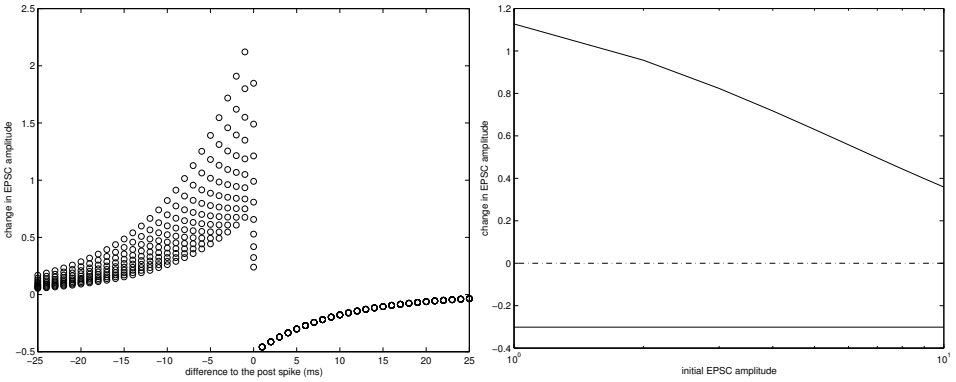

FIG. 3.1.9 – *Modèle génératif de la STDP.* Comme dans Fig. 3.1.8, une synapse est excitée de façon répétée par un PA pré-synaptique et un PA post-synaptique dont le délai relatif est constant : on utilise alors le modèle génératif de plasticité [Perrinet et Samuelides, 2000] décrit dans le texte avec $\bar{g}_r = 0,1$, $\tau_{\texttt{pre}} = 15\text{ms}$, $\tau_{\texttt{post}} = 25\text{ms}$, $\alpha_{\texttt{pre}} = 0,5$, $\alpha_{\texttt{post}} = 0,5$, $A^{\texttt{actif}}_{\texttt{pre}} = 2$ et $A^{\texttt{actif}}_{\texttt{post}} = 0,5$. **(A)** Variation de l'efficacité de la synapse en fonction du délai du PA pré-synaptique relativement au PA post-synaptique. **(B)** Variation de l'efficacité relative en fonction de l'amplitude initiale de l'efficacité.

remarque aussi que comme décrit dans Fig. 3.1.8-(B), cette expression fait en sorte que la dépression soit divisive et la potentiation additive dans le cas particulier où $\bar{g}_r$ est faible Fig. 3.1.9-(B). Cette règle, paramétrée par les paramètres de mécanismes synaptiques à court terme peuvent alors, avec un jeu différent de paramètres, prendre des formes très diverses comme par exemple une règle anti-hebienne.

Il est alors intéressant de simuler le modèle complet de neurone impulsionnel que nous avons bâti et qui comprend alors les mécanismes membranaires et les mécanismes synaptiques à court et long terme. Son comportement se rapproche sur de nombreux points aux neurones biologiques. Les poids sont modifiés de telle façon à ce que la fréquence de décharge se stabilise quelle que soit la fréquence de décharge des PAs en entrée, et donc de ce modèle peut émerger un processus homéostatique de régulation [Kempter et al., 2000; Chechik et Tishby, 2000]. Cette régle est particulièrement adaptée par des codes se basant sur la synchronie des PAs et a souvent été étudiée dans ce cas.

Détection de la cohérence

Une relecture de la règle de Hebb [1949] montre qu'elle prend vraiment son sens dans des assemblées de neurones. Dans ce cadre, nous allons utiliser une approche fonctionnelle en dérivant cette règle d'un coût qui quantifie sa réussite. Une méthode de descente de gradient nous fournira alors une règle d'apprentissage sur cette assemblée que nous rapprocherons à la méthode précédente [Perrinet et Samuelides, 2002a]. Nous étudierons une assemblée de neurones telles que les cellules ganglionnaires de type "B" de la rétine du crapaud qui n'ont pas d'arbre dendritique mais qui possèdent une quarantaine de boutons synaptiques sur leurs soma. Nous modéliserons alors des cellules de ce type qui partagent les mêmes entrées, tout en utilisant le modèle SRM. Dans un premier temps, l'entrée est constituée de *motifs synchrones*, définis par les sous-ensembles de synapses qui seront activées au même moment. Le but de l'assemblée est alors que les neurones s'adaptent de façon autonome à l'entrée synaptique pour finalement répondre à un motif et à un seul, on aura détecté la cohérence du signal impulsionnel. Notre but est de trouver pour cette assemblée une règle non-supervisée d'apprentissage qui converge vers ce comportement quelle que soit l'ensemble des motifs.
Formellement, donnons nous une assemblée de $N_{\mathtt{rec}}$ neurones i partageant $N_{\mathtt{em}}$ entrées j et que l'information qui arrive sur cette assemblée est constituée de $N_{\mathtt{sig}}$ signaux synchrones différents, indexés par k et caractérisés par le sous-ensemble $\mathcal{M}_k$ des synapses qui sont activées de façon "presque" synchrone (le bruit de synchronie est fixé à environ $\tau_{\mathtt{sync}} \sim 2\,\mathrm{ms}$). Ces neurones ne sont pas reliés par des connexions latérales et leur évolution sont donc découplées. On peut, pour une matrice des poids synaptiques donnée, simuler la dynamique de ces neurones quand l'on présente ces motifs noyés dans un bruit de fond (modélisé par un processus de Poisson) et fixer les paramètres pour qu'ils émettent des impulsions. Dans notre simulation, les neurones répondent donc aux signaux synchrones qui sont choisis au hasard et présentés aléatoirement au réseau suivant un processus de Poisson de faible fréquence (4Hz, soit une présentation en moyenne toutes les 250 ms).
Nous allons définir un coût de type hebbien dans le cadre des neurones im-

pulsionnels en associant une adaptation synaptique à chaque tir d'un neurone récepteur, qui sera alors noté le neurone "gagnant" i^n (et n sera donc le temps d'apprentissage). Le coût est défini comme la somme de la divergence des potentiels membranaires d'une valeur souhaitée V^*. Cette valeur souhaitée est fixée de telle sorte que le neurone gagnant corresponde à une valeur haute (comme le potentiel $E_{\texttt{exc}}$) alors que les neurones qui ont perdu correspondent à une valeur basse (comme $E_{\texttt{inh}}$), soit une réponse de type *Winner Take All*. On ajoute à ce coût des termes de régularisation qui correspondront à un évaluation de l'énergie dépensée en courants synaptiques et aussi à une évaluation des ressources utilisées (le nombre de canaux synaptiques nécessaires). Le coût est défini à l'instant où le neurone gagnant tire par (on a écrit V_i pour $V_i(t^n)$) :

$$2.E = \sum_{1\leq i\leq N_{\texttt{rec}}} (V_i - V_i^*)^2 + \alpha \sum_{1\leq i\leq N_{\texttt{rec}}} (\frac{\partial V_i}{\partial t})^2 + \beta. \sum_{\substack{1\leq i\leq N_{\texttt{rec}} \\ 1\leq j\leq N_{\texttt{em}}}} {w_{ij}}^2 \qquad (3.1.18)$$

les potentiels V_i^* étant

$$\left\{ \begin{array}{lcll} V_{i^n}^* & = & E_{\texttt{exc}} & \\ V_i^* & = & E_{\texttt{inh}} \qquad \text{pour} \qquad & i \neq i^n \end{array} \right. \qquad (3.1.19)$$

On règle α et β comme des paramètres régularisateurs. On peut alors calculer le gradient de ce coût et le simplifier grâce à Éq. 3.1.13 :

$$\frac{\partial E}{\partial w_{ij}} = (V_i - V_i^*).\frac{\partial V_i}{\partial w_{ij}} + \alpha.\frac{\partial V_i}{\partial t}.\frac{\partial^2 V_i}{\partial w_{ij}\partial t} + \beta.w_{ij} \qquad (3.1.20)$$

$$= (V_i - V_i^*).p_j + \alpha.\frac{\partial V_i}{\partial t}.\frac{dp_j}{dt} + \beta.w_{ij} \qquad (3.1.21)$$

où $p_j(t)$ est la réponse à chaque entrée impulsionnelle S_j (définie par ses temps de tir $t_j^m \in \Gamma_j$) par des synapses dont les poids sont normalisés à 1 :

$$p_j(t) = (S_j * \varepsilon)(t) = \sum\nolimits_{t_j^m \in \Gamma_j} \varepsilon(t - t_j^m) \qquad (3.1.22)$$

On peut alors induire un apprentissage en minimisant progressivement le coût à chaque PA (on compte le temps d'apprentissage n par les événements post-synaptiques) par une méthode de descente de gradient [Amari, 1998]. Il

vient alors (on a écrit p_j pour $p_j(t^n)$) :

$$w_{ij}^{n+1} = w_{ij}^n - \gamma_n.\frac{\partial E}{\partial w_{ij}} \quad (3.1.23)$$

$$= (1-\beta.\gamma_n).w_{ij}^n + \gamma_n.(V_i^* - V_i).p_j + \alpha.\gamma_n.\frac{\partial V_i}{\partial t}.\frac{\partial p_j}{\partial t} \quad (3.1.24)$$

avec γ_n satisfaisant la condition de convergence pour les algorithmes stochastiques $\sum_{n=1...\infty} \gamma_n \rightarrow \infty$ et $\sum_{n=1...\infty} {\gamma_n}^2 < \infty$.
Finalement, on en déduit une règle de modification de l'efficacité synaptique que l'on peut rapprocher des formulations de Kempter et al. [1999] induites par les expériences neurophysiologiques, mais qui est généralisée à une assemblée de neurones et tirée d'une formulation fonctionnelle.

$$w_{ij}^{n+1} = (1-\gamma_n).w_{ij}^n + \alpha_n.(V_i^* - V_i).p_j + \beta_n.(\sum w_{ij}.\frac{dp_j}{dt})\frac{dp_j}{dt} \quad (3.1.25)$$

Cette équation est alors facilement implantable en utilisant des algorithmes similaires à la méthodologie du SRM.
Tout d'abord, on remarque que si l'on réplique les conditions de l'expérience de Bi et Poo [1998] sur une synapse et un seul neurone, alors le comportement de l'efficacité synaptique sera similaire (voir Fig. 3.1.10). Toutefois, la forme de la fenêtre d'apprentissage sera différente et liée à la forme de ε et aux statistiques des PAs. On remarque aussi que les neurones non associés (c'est à dire ceux qui ne sont pas gagnants et pour lesquels $V_i^* - V_i = E_{\texttt{inh}} - V_i < 0$) apprendront de manière anti-hebbienne. Les résultats de cette règle montrent alors, comme pour le neurone seul, que la norme des poids converge vers une valeur dépendant seulement de l'activité. Cette règle, suivant un principe homéostatique, implique que la fréquence de décharge est relativement constante et que l'activité reste irrégulière.

Aussi, les simulations montrent que la distribution des poids des différents neurones devient bi-modale, comme il a été observé dans certains neurones du SNC [Song et al., 2000]. Une synapse pourra donc être considérée soit activée ou soit désactivée et on observe que cette force est modulée par la distribution des signaux $\mathcal{M}_k$, le poids d'une synapse activée chez tous les signaux convergera vers une valeur moins forte que pour une synapse parti-

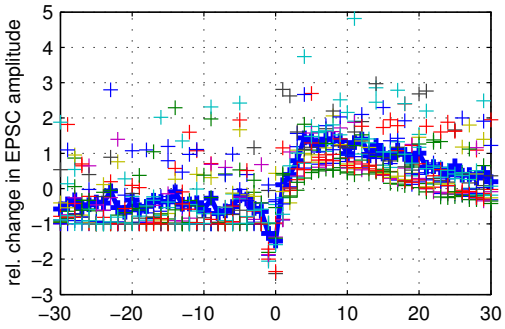

FIG. 3.1.10 – *Détection non-supervisé de cohérence et STDP.* En reproduisant les conditions de l'expérience de Bi et Poo [1998], on a tracé pour chaque pas de l'apprentissage (donc chaque tir du neurone récepteur) le changement relatif de l'efficacité en fonction du délai du neurone pré-synaptique. Le modèle observe un comportement qui peut se rapprocher à la STDP (voir Fig. 3.1.8-(A)), la forme de la fenêtre se rapprochant plus d'un PPS. Toutefois, la règle de STDP appliquée à une assemblée de neurones n'est pas suffisante pour permettre la détection de la cohérence des impulsions.

culière à un seul signal. Cette efficacité est aussi d'autant plus forte que la synapse est peu activée, de sorte qu'une synapse qui serait active dans tous les motifs serait considérée comme non-significative. L'inhibition correspondant aux perdants de la compétition permet aussi de faire en sorte qu'un neurone et un seul tire pour chaque signal. Finalement, le système converge de façon non-supervisée pour faire correspondre un motif à chaque neurone (voir Fig. 3.1.11).

Une extension de ce modèle consiste à influencer la convergence par le profil de V_i^* suivant les fibres voisines du gagnant* pour obtenir à l'image des cartes associatives de Kohonen [1982] une correspondance spatiale à des associations complexes, comme cela a été abordé avec les "espaces de visages". Cette expérience montre que cette méthode peut fournir des moyens simples de détection d'entrées synchrones. Il faut alors étendre ce système à des réseaux et des principes plus complets et particulièrement à des motifs asynchrones.

*Par exemple, $V_i^* = \exp(-\|i - i_n\|/T)$, où $\|.\|$ correspond à une distance et T à une "température"

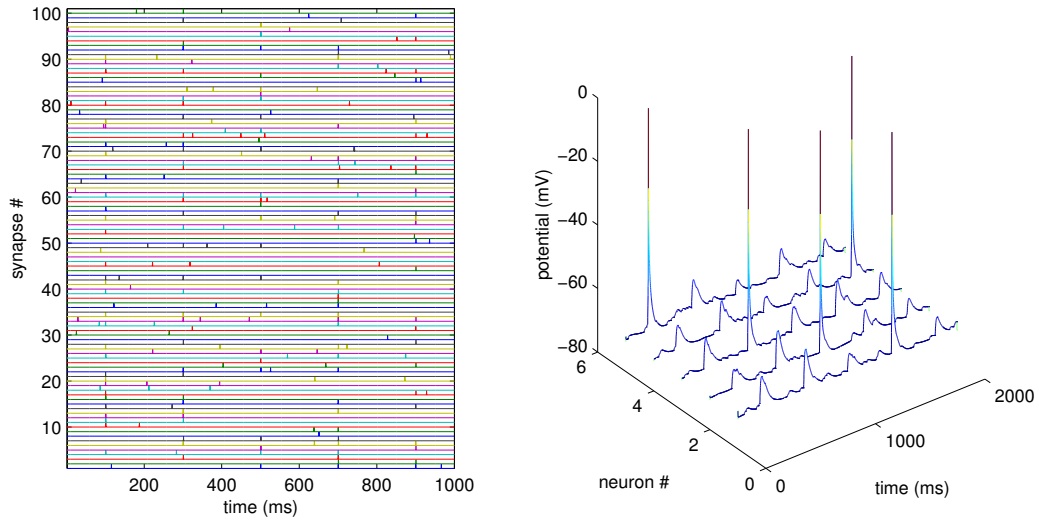

FIG. 3.1.11 – *Détection de la cohérence par STDP.* **(Gauche)** L'entrée est un flux constant de PAs, que l'on peut caractériser comme un processus ponctuel de Poisson de moyenne $1.5\mathrm{Hz}$ pour chaque fibre et sur lequel on ajoute 5 signaux synchronisés différents (ici à $t = 100ms, 300ms, 500ms, 700ms, 900ms$). **(Droite)** La couche neurale consiste en 5 neurones qui apprennent de façon non-supervisée suivant la règle décrite dans le texte (voir Éq. 3.1.25). Après environ $100\mathrm{s}$ (en temps du réseau simulé), le réseau a convergé et répond aux différents signaux et seulement à un de ceux-ci : le réseau s'est donc adapté à l'entrée pour détecter les signaux cohérents.

3.1.3 Codage impulsionnel par la latence de la première impulsion

Les modèles que nous avons présentés dans les sections précédentes permettent de répliquer des aspects fondamentaux du traitement neural et de sa faculté d'adaptation. Mais le fonctionnement du code neuronal reste difficile à découvrir. Nous allons essayer d'extraire les bases d'un code impulsionnel qui permette de coder par la latence de tir de ces impulsions l'information lumineuse. Nous allons donc étudier les principes pour coder et décoder par la latence mais aussi comment adapter le réseau de neurones pour maximiser l'efficacité du traitement.

Le code doit transmettre une information

Tout d'abord, à l'image du flux impulsionnel dans le nerf optique, une première fonction du code est de pouvoir transmettre une information de façon fiable et robuste. Même s'il est aberrant d'imaginer qu'il existe à une étape des chemins de la vision une reconstruction "analogique" de l'image, il est établi que le rôle de certaines commissures nerveuses (comme le nerf optique) est de transmettre sans transformation une information. Nous étudierons alors la reconstruction comme une borne supérieure de la qualité de la transmission de l'information.
Il semble alors que pour des signaux dans la gamme dynamique des signaux naturels saillants, la réponse des neurones est relativement robuste au bruit*. En effet, des enregistrements *in vitro* de neurones corticaux montrent que, suite à la présentation d'un stimulus, ces neurones sont capables de produire des décharges très reproductibles [Mainen et Sejnowski, 1996; de Ruyter van Steveninck et al., 1997]. En particulier, la latence de décharge de ces neurones peut être très précise, de l'ordre de la milliseconde [Bair et Koch, 1996], et la latence de chaque impulsion peut donc porter une grande quantité d'information. En pratique, on observe dans le cortex primaire du singe que la

*En fait, certain processus, comme la libération synaptique peuvent se révéler hautement aléatoires. Toutefois nous considérerons que ce "bruit" est négligeable pour des neurones fonctionnels dans la gamme des signaux auxquels ils sont adaptés.

direction locale d'une barre par les neurones est codée par la différence de latence depuis une latence minimale correspondant à une direction préférée [Celebrini et al., 1993] : on en déduit qu'un code impulsionnel de traitement rapide de l'information visuelle (voir Sec. 2.3) peut être implanté en utilisant la latence de décharge de la première impulsion. Cette hypothèse a été proposée par Thorpe [1990] et implique une dualité dans ce codage. En effet, la valeur excitatrice analogique pourra être restituée si l'on connaît sa latence exacte, mais de plus la propagation est progressive et donc l'ordre des tirs va influencer le résultat du calcul. En codant l'information de façon parallèle et asynchrone, les neurones peuvent implanter des calculs en utilisant resp. la position et la latence des PAs.
Cette hypothèse est appuyée par les modèles neurophysiologiques (voir Sec. 3.1) que nous avons présentés puisque nous avons vu puisque, par exemple pour le modèle Intègre-et-Tire, il existe pour une excitation constante une relation bijective (monotone décroissante) entre l'amplitude de l'entrée et la latence de décharge. Toutefois, à la différence des calculateurs synchrones, il n'existe pas dans le SNC d'horloge absolue pour chaque neurone et le décodage de cette information ne peut être que relatif à une activité d'ensemble, comme les activités γ qui semblent réguler l'activité de nombreux neurones. La coopération entre les différents neurones dans le temps —implicite à un modèle utilisant l'asynchronie du signal— est donc nécessaire pour distinguer un signal préféré d'un afflux massif d'impulsions.

Détecter un front asynchrone d'impulsions

Grâce aux modèles mathématiques que nous avons introduits, par exemple le modèle SRM, des configurations synaptiques différentes peuvent permettre de catégoriser des informations et que ces catégories pouvaient être apprises de façon non-supervisée. Étudions le comportement d'un neurone au repos sur lequel converge une volée de PAs. Ainsi, si l'on modélise ce neurone i par ses poids w_{ij} et que l'on ne considère que les premiers PAs arrivent sur la synapse à des latences t_{ij} alors on pourra noter l'accroissement ΔV_i d'acti-

vité[†] dans la phase linéaire du potentiel post-synaptique (c'est-à-dire avant environ 5ms) :

$$\Delta V_i(t) = \sum_j w_{ij}.\varepsilon(t - t_{ij}) \tag{3.1.26}$$

$$\sim t.\sum_j [w_{ij}.\Delta(t - t_{ij})] \tag{3.1.27}$$

On rappelle que Δ désigne la fonction créneau, voir Sec. 3.1.1, Éq. 3.1.9. On en déduit que comme le neurone de McCullough, un codage temporel par la latence peut permettre de calculer une somme pondérée et donc de détecter une séquence particulière de latences, même si ce modèle répondra prioritairement au flux total d'impulsions et donc répondra particulièrement lors d'un afflux massif d'impulsions sur les synapses excitatrices. Toutefois, ce modèle n'est pas encore satisfaisant car il répondrait indifféremment à un afflux massif de PAs et qu'il donne le même poids aux impulsions quel que soient leur ordre d'arrivée.

De façon plus générale, on peut alors montrer que les neurones à impulsions peuvent calculer un nombre arbitraire de fonctions [Maass et Zador, 1999]. Nous devons donc étendre les signaux cohérents à des signaux asynchrones qui à la différence de signaux synchrones vont pouvoir transmettre une valeur modulée, comme cela est produit dans une assemblée de neurones. Dans un modèle de V1, par exemple, on pourra donc imaginer que la similitude la plus forte avec une orientation préférée sera transmise avant les autres et que le "calcul neural" correspondra plutôt à un flux d'information. Dans cette analogie, les neurones jouent plutôt le rôle de relais s'activant d'autant plus rapidement que le flux qui les touche correspond à une classe préférée de motifs.

Ces premiers résultats suggèrent alors des principes "macroscopiques" de calcul non-linéaires : une réponse à un flux incohérent doit être codé par une sortie régulière alors qu'une entrée cohérente (donc porteuse d'information) doit être codée par une sortie irrégulière [Zador, 1998]. Ces résultats suggèrent aussi que l'information la plus importante est propagée en premier

[†]Ce calcul est dérivé d'une formulation similaire de Maass et Zador [1999].

et donc qu'un code efficace doit pouvoir prendre en compte l'ordre d'arrivée des impulsions [Thorpe, 1990].

Adaptation du codage temporel

Considérons la figure Fig. 3.1.12-(A) : si on essaye d'évaluer subjectivement les contrastes entre les niveaux de gris (définis ici quantitativement par la densité de noir : fond clair, 90%, gris 50% et fond noir, 10%), le gris apparait plus contrasté sur fond sombre alors que des différences de luminance indiqueraient un contraste équivalent. En fait, comme d'autres sensations régies par la *loi de Stevens*, notre perception de l'intensité lumineuse, la *luminosité*, est plutôt sensible à des *ratios* de luminance[‡]. Ce point révéle une donnée physique souvent éclipsée en traitement de l'image et dont les conséquences peuvent être cruciales. En effet, la réponse neuronale à une luminance varie selon une loi de puissance dont l'exposant permet de faire varier l'"étalement" de l'histogramme moyen des luminosités. En effet, on observe que dans les images naturelles les luminances basses sont plus fréquentes (voir Fig. 3.1.12-(B), courbe pleine aux basses valeurs) et que l'on doit la transformer en une luminosité suivant un paramètre tel que la loi de probabilité résultante corresponde à une meilleure répartition des valeurs, c'est à dire que les réponses soient mieux différentiées pour des intensités différentes. Ce paramètre est appelé le facteur γ [Poynton, 1999].
On peut écrire formellement ce principe intuitif en définissant l'*entropie* d'un code et en déterminant alors la transformation optimale de la luminance pour qu'elle représente au mieux l'information. En effet, l'entropie est la moyenne de l'information $I(w) = -\log_2(P(w))$ portée par les différents messages w (de probabilité $P(w)$), soit

$$H(w) = \langle I(w) \rangle = -\sum\nolimits_w P(w).\log_2(P(w)) \tag{3.1.28}$$

[‡]La luminosité est définie par un standard comme $L = 116(\frac{Y}{Y_n})^{\frac{1}{3}} - 16$ [Central Bureau of the Commission Internationale de L'Éclairage (Vienna, Austria), 1986], où Y et Y_n sont respectivement la luminance (c'est-à-dire une mesure du flux lumineux) et la luminance standard du blanc.

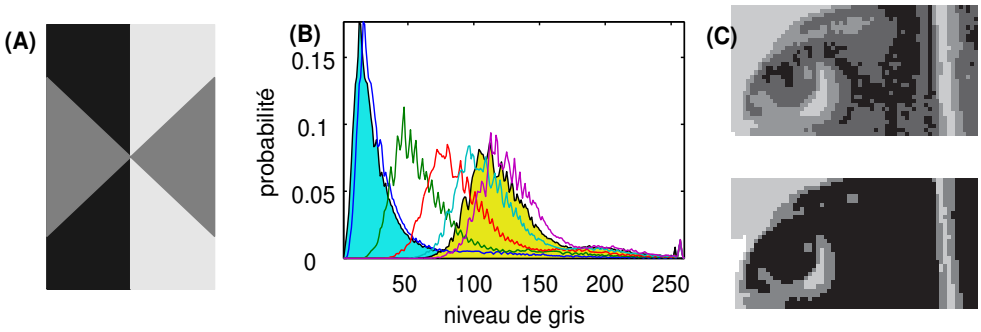

FIG. 3.1.12 – *Luminance et luminosité.* **(A)** Effet de contraste : cette figure présente un "papillon" horizontal de luminance 50% (le blanc correspondant à 100%, le noir à 0%) par rapport à un fond clair (luminance 90%) ou symétriquement foncé (luminance 10%). Pourtant, le contraste subjectif semble plus fort à droite : nous percevons la luminosité, c'est-à-dire des ratios de luminance (Pour un γ non standard de l'impression ou de l'affichage de cette figure, l'effet de contraste peut lui-même être distordu, on vérifiera sur une impression noir et blanc tramée que la surface d'encre noire pour la luminance 50 est équilibrée avec la surface non couverte de papier blanc.), de telle sorte sue la luminance n'est pas perçue identiquement à gauche et à droite. **(B)** Histogramme moyen des luminosités d'images (images calibrées tirées de la base de donnée mise à disposition par van Hateren [1993]) corrigées par γ croissants. La courbe pleine pour les faibles luminosités correspond à l'histogramme moyen des luminances ($\gamma = 1$). L'entropie maximale est réalisée pour $\gamma = 2.8$ (courbe pleine centrale). **(C)** Effet de crénelage sur des zones sombres après une diminution du nombre de bit par pixel : (Haut) codage par luminance ($\gamma = 2.8$) et (Bas) par luminosité ($\gamma = 1$). Le codage par luminosité peut faire perdre de l'information, comme ici par exemple dans le gradient sombre qui entoure ici l'œil de Lena.

Une stratégie est alors de maximiser l'entropie du code (*i.e.* l'information moyenne des messages utilisés), et en pratique ce processus revient à trouver la densité de probabilité de la valeur transformée qui soit la plus étalée. Cela revient donc à une égalisation moyenne de l'histogramme d'un ensemble d'images. En pratique, nous avons calculé pour des images naturelles non calibrées (les valeurs des pixels correspondant aux luminances) et tirées de la base de données utilisée dans van Hateren et Snippe [2001] pour calculer l'entropie de la distribution de la luminance et de ses transformations par des lois logarithmiques de bases différentes. Le facteur optimal pour la base de donnée est alors égal à environ $\gamma \sim 2.8$ (voir Fig. 3.1.12-(B), courbe pleine centrale) et la luminosité va donc correspondre à la luminance transformée par cette loi de puissance.
Aussi la technologie s'est de même adaptée, et les standards d'imagerie utilisent le plus souvent une loi en puissance pour coder ou décoder en luminance la luminosité. Malheureusement, les standards se sont multipliés et ce facteur est mal connu[§], rendant le risque fort qu'une image transformée suivant des cascades de processus par des personnes et des medium différents perde progressivement de l'information (voir Fig. 3.1.12-(C)).
On déduit finalement que l'on peut adapter la gamme dynamique de réponse d'un neurone pour qu'il représente le mieux l'information. L'utilisation d'une mesure statistique, comme l'entropie, permet alors de quantifier plusieurs stratégies et de trouver les meilleurs paramètres du codage. Ce résultat est d'autant plus important dans notre cadre car l'information en plus d'être codée temporellement est propagée dans un ordre précis. Il est donc primordial pour un code temporel que l'information qui est transmise en avant soit bien l'information qui porte le plus d'information. On pourra ainsi définir une coopération entre des neurones qui pourront ouvrir à l'implantation d'un

[§]Le facteur utilisé par le système de télévision NTSC utilise $\gamma = 2.2$ et PAL $\gamma = 2.8$; pour les moniteurs d'ordinateurs, WINDOWS© pose $\gamma = 2.2$ et APPLE© permet de le régler. Aussi, le problème s'étend à l'impression puisque par exemple la relation entre la quantité d'encre sur le papier et la luminosité introduit une nouvelle non-linéarité qui peut être relié à un nouveau facteur γ pour chaque imprimante. Ceci impose aussi le format PNG —distribué sous licence GPL sur http ://www.libpng.org/ et compatible maintenant avec la majorité des logiciels— comme une référence en particulier pour le traitement de l'image car il prend explicitement en charge la valeur du γ.

code impulsionnel.

3.2 Modèle dynamique de la rétine

La seconde découverte majeure en neurosciences, aussi récompensée par un prix Nobel, fut la découverte par Hubel et Wiesel [1959] des champs récepteurs sélectifs à l'orientation dans l'aire V1 du cortex du chat. Les modèles basés sur ces observations constituent avec le modèle Hodgkin-Huxley deux "piliers" de la neurosciences et montrent la divergence qu'il existe dans cette discipline au niveau de description au-dessus ou au-dessous de la cellule. Nous allons prendre l'exemple de la rétine pour décrire un modèle complet de transformation impulsionnelle qui puisse s'intégrer dans un modèle neural plus complet tout en restant compatible avec le traitement rapide de l'information visuelle.

3.2.1 Codage multi-échelle de l'information rétinienne

Nous avons décrit la rétine dans la Sec. 2.2.1 comme un convertisseur d'une image analogique en un front de PAs et nous allons essayer ici de proposer un modèle de cette transformée tout en exploitant les résultats des sections précédentes. Cette section va reprendre le modèle présenté dans Van Rullen et Thorpe [2001] dans un cadre mathématique plus général. La fonction de ce modèle va être de transmettre par un code impulsionnel le signal à travers le canal informationnel du nerf optique. Nous allons alors formaliser notre modèle de rétine comme un traitement parallèle sur des couches correspondant à la transformée linéaire de l'image en une représentation compacte, sa transformée en PAs, puis enfin le récepteur, la couche de neurones du CGL qui permet —même si ce n'est pas biologiquement réaliste— la reconstruction théorique de l'image (voir Fig. 3.2.1).

Architecture de la transformée linéaire

Comme l'ont suggéré les mesures neurophysiologiques de Rodieck [1965], nous allons modéliser la transformation de l'image par la rétine comme une

transformation linéaire. Celle-ci est caractérisée sur une couche de cellules ganglionnaires i, définies par leurs vecteurs de poids synaptiques ϕ_i sur leurs champs récepteurs $\mathcal{R}_i$ (voir Fig. 3.2.1-left). Nous écrirons de façon générale le potentiel C_i induit par la présentation de l'image comme le produit scalaire [Atick et Redlich, 1992] :

$$C_i =< I, \phi_i >= \sum\nolimits_{\vec{l}\in\mathcal{R}_i} I(\vec{l}).\phi_i(\vec{l}) \tag{3.2.1}$$

où $I(\vec{l})$ est l'intensité lumineuse à la position $\vec{l}$ (en informatique, resp. la valeur et position du *pixel*). D'après Field [1994], nous définirons les filtres d'après des gaussiennes G_σ comme des *Différences De Gaussiennes* (filtres DDG , voir [Mallat, 1998, pp. 77]),

$$DDG(\vec{l}) = 9 * G_\sigma(\vec{l}) - G_{3*\sigma}(\vec{l}) \tag{3.2.2}$$

avec

$$G_\sigma(\vec{l}) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{\|\vec{l}\|^2}{2\sigma}) \tag{3.2.3}$$

qui approchent les "filtres" observés dans la rétine biologique [Enroth-Cugell et Robson, 1966]. On pourra aussi utiliser la dérivée seconde de la fonction gaussienne (la *Laplacienne de Gaussienne*, LDG), qui est utilisée en traitement de l'image comme un filtre de détection de contours*.

$$LDG(\vec{l}) = \Delta G_\sigma(\vec{l}) \tag{3.2.4}$$

$$= \frac{2}{\sqrt{3\sigma\sqrt{\pi}}}(2 - \|\vec{l}\|^2)\exp(-\frac{\|\vec{l}\|^2}{2}) \tag{3.2.5}$$

$$\approx (1.2^2) * G_\sigma(\vec{l}) - G_{1.2*\sigma}(\vec{l}) \tag{3.2.6}$$

La dernière équation (Éq. 3.2.6) approchant, comme l'avait suggéré Marr [1980], ce filtre par une différence de gaussiennes (voir Fig. 3.2.2). Dans ces définitions σ correspond à un facteur d'échelle de la largeur du filtre, et par une simple translation du vecteur de poids, on pourra donc définir un filtre quelconque i de même norme par sa position et sa taille $i = \{\sigma, \vec{\lambda}\}$ grâce à

*En particulier, les passages par zéro de cette transformée indique un maximum du gradient local à l'échelle σ, soit une localisation d'un contour local

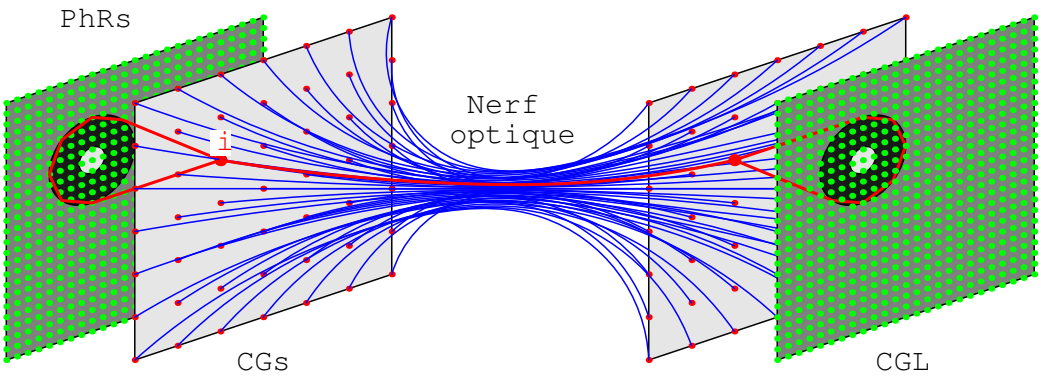

FIG. 3.2.1 – *Architecture du modèle dynamique de rétine.* Le système est composé d'une couche de photo-récepteurs (PhRs) dont les activités se combinent linéairement sur la couche de cellules ganglionnaires (CGs) qui produisent une volée de PAs. Ceux-ci sont transmis sans bruit par le nerf optique pour induire une activité dans le corps genouillé latéral. Pour quantifier la qualité de la transmission, on imagine -contre ce qui est observé dans le système visuel- que l'image est reconstituée au CGL. Un neurone i est souligné en rouge, laissant apparaître son champ récepteur centre-ON, son axone dans le nerf optique, sa synapse à un neurone du CGL et le champ émetteur correspondant dans la couche du CGL.

$\phi_\sigma(\vec{l}) = \frac{1}{\sqrt{\sigma}}\phi(\frac{\vec{l}}{\sigma})$ et à $\phi_i(\vec{l}) = \phi_\sigma(\vec{l} - \vec{\lambda})$.

Si l'on considère alors des cellules ganglionnaires dont les centres sont placés sur des grilles rectangulaires régulières dont l'échantillonnage s'accroît géométriquement en même temps que la taille des champs récepteurs associés, les activités calculées vont correspondre, par définition, aux coefficients d'une *transformée en ondelettes* [Mallat, 1998]. Plus particulièrement, comme suggéré par [Van Rullen et Thorpe, 2001], on peut choisir une grille *dyadique*, c'est-à-dire où les échelles grandissent par un facteur de 2 sur chaque dimension. Finalement, les activités des cellules ganglionnaires calculées pour ces différents filtres vont simplement correspondre à une transformée en ondelettes de l'image sur une grille dyadique.

À l'échelle 1, si on choisit les filtres DDG, on prendra alors $\sigma_1 = .5$ pour que les filtres recouvrent correctement l'image[†] et on aura donc à l'échelle s

[†]Cette hypothèse n'est pas explicite dans l'article original mais peut être retrouvée en maximisant une condition que nous verrons plus bas (voir Éq. 3.2.8).

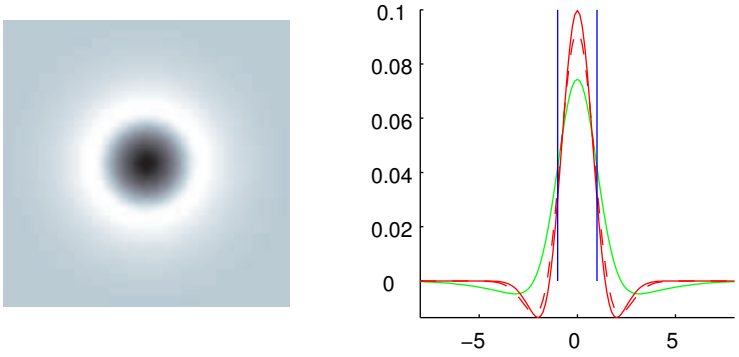

FIG. 3.2.2 – *Filtres linéaires de la rétine.* (**Gauche**) Filtre annulaire DDG (voir Éq. 3.2.2). (**Droite**) Coupe radiale des profils du filtre LDG et de son approximation par une différence de gaussiennes (traits pointillés), puis du filtre DDG de Field [1994], pour un même σ. Les traits verticaux marquent l'échelle $\sigma = 1$ (et donne donc une unité de référence en abscisse)

(jusqu'à l'échelle maximale $s_{\max}$), $\sigma_s = \sigma_1.2^{s-1}$. Le calcul n'est effectué pour chaque taille s que sur la grille sous échantillonnée $\mathcal{D}_s$ d'un facteur 2^{s-1}, c'est-à-dire sur les points de la forme $(x, y) = (2^{s-1}.i, 2^{s-1}.j)$ avec (i, j) entiers naturels. Ce processus conduit à la création d'une grille globale multicouche $\mathcal{D} = \bigcup_{1 \leq s \leq s_{\max}} \mathcal{D}_s$ des points de calcul en fonction de la taille. Le nombre total de neurones sera supérieur au nombre de pixels d'un facteur multiplicatif de $\chi = 1 + (1/2)^2 + \ldots + (1/2)^{(2*s_{\max})} = \sum_i (1/2)^{2*(i-1)} = \frac{1-(1/2)^{s_{\max}}}{1-(1/2)^2}$ soit environ 4/3. Par souci de réalisme, on peut choisir de séparer deux populations ON et OFF de filtres et étudier la décomposition en coefficients rectifiés. Ceci donne en comptant les polarités un nombre total de 8/3 le nombre de points dans l'image, mais n'apporte aucune information supplémentaire au "calcul" réalisé dans cette architecture par la rétine[‡].

[‡]On notera que cette symétrie ON / OFF est tributaire de la symétrie de la loi de probabilité des contrastes. Celle-ci n'est réalisée que grâce aux principes qui transforment la luminance et par la symétrie de la loi de probabilité de la luminosité qui résulte de cette transformation.

Reconstruction idéale de l'image

La théorie des ondelettes permet de prouver que l'on peut reconstruire l'image à partir des coefficients de la transformée si la transformée définit un *cadre* (ou *frame* en anglais). C'est le cas en général pour une classe detranformée en ondelettes et la formule de Calderòn permet de reconstruire de façon générale l'image depuis les coefficients de la transformée. Cette architecture particulière permet d'écrire une formule plus simple de reconstruction.

Théorème. *Soit I une image et $C_i =< I, \phi_i >$ pour $i = \{\sigma, \vec{\lambda}\}$ les coefficients correspondants à la transformée en ondelettes dyadique. Si on note $\overline{\phi_\sigma}$ le complexe conjugué du symétrique de ϕ_σ (*i.e. $\overline{\phi_\sigma} = \phi_\sigma^*(-.)$*) et*

$$K = \sum_{\sigma=\sigma_1.2^{(s-1)} 1 \leq s \leq N} \frac{1}{\sigma^2}(\overline{\phi_\sigma} * \phi_\sigma)$$

Alors, la reconstruction

$$I_{rec}(k,l) = \sum\nolimits_{i \in \mathcal{D}} C_i.\phi_i(k,l)$$

sera telle que $I_{rec} = I * K$.

Démonstration. Étudions d'abord l'image des coefficients à une taille σ. Les coefficients forment sur les points (x, y) d'échantillonnage de $\mathcal{D}_\sigma$ une image C_σ telle que

$$C_\sigma(x,y) =< I, \phi_{(x,y,\sigma)} >= \sum\nolimits_{(k,l) \in \mathcal{R}_{(x,y,s)}} I(k+x, l+y).\phi_{(x,y,\sigma)}(k,l)$$

Avec $D_\sigma = \delta_{\mathcal{D}_\sigma}$ la matrice nulle sauf sur les points de la grille de sous-échantillonnage à la taille σ. Alors cette image peut s'écrire comme une convolution avec les symétriques des filtres, échantillonnée aux points de $\mathcal{D}_\sigma$.

$$C_\sigma = (I * \overline{\phi_\sigma}).D_\sigma$$

On pourra alors écrire la reconstruction c_σ restreinte aux coefficients C_σ (et

donc limitée à l'échelle σ).

$$c_\sigma = \sum\nolimits_{i=(x,y,\sigma)\in\mathcal{D}_\sigma} C_\sigma(x,y).\phi_i$$

soit

$$c_\sigma = C_\sigma * \phi_\sigma = ((I * \overline{\phi_\sigma}).D_\sigma) * \phi_\sigma$$

soit dans l'espace de Fourier,

$$\begin{aligned}\widehat{c_\sigma} &= ((\widehat{I}.\widehat{\overline{\phi_\sigma}}) * \widehat{D_\sigma}.)\widehat{\phi_\sigma} \\ &= ((\widehat{I}.\widehat{\phi_\sigma}^*) * \widehat{D_\sigma}).\widehat{\phi_\sigma}\end{aligned}$$

or, comme nous sommes en 2D et que les filtres sont réels,

$$\widehat{\phi_\sigma}(\omega_x,\omega_y) = \sigma.\widehat{\phi}(\sigma.\omega_x,\sigma.\omega_y)$$

Aussi, comme $D_\sigma = \sum_{(k,l)\in\mathbb{Z}^2}\delta_{(\sigma.k,\sigma.l)}$, d'après la formule de Poisson (voir [Mallat, 1998, p. 29]) adaptée au cas bidimensionnel,

$$\widehat{D_\sigma} = \frac{4\pi^2}{\sigma^2}\sum_{(k,l)\in\mathbb{Z}^2}\delta_{(\frac{2\pi k}{\sigma},\frac{2\pi l}{\sigma})}$$

finalement,

$$\begin{aligned}\widehat{c_\sigma}(\omega_x,\omega_y) &= 4\pi^2 \sum_{(k,l)\in\mathbb{Z}^2}[\widehat{I}(\omega_x-\frac{2\pi k}{\sigma},\omega_y-\frac{2\pi l}{\sigma}). \\ &\quad \widehat{\phi}^*(\sigma.\omega_x-2\pi k,\sigma.\omega_y-2\pi l).\widehat{\phi}(\sigma.\omega_x,\sigma.\omega_y)]\end{aligned}$$

Or si on considère que $\widehat{\phi}$ est de support compact compris dans $[-\pi,\pi]\times[-\pi,\pi]$, alors pour $\sigma \geq 1$, $\widehat{\phi_\sigma}$ l'est aussi et les termes de la somme pour $(k,l)\neq(0,0)$ sont donc nuls. On peut donc simplifier la somme et trouver $\widehat{c_\sigma}$ par

$$\widehat{c_\sigma}(\omega_x,\omega_y) = 4\pi^2.\widehat{I}(\omega_x,\omega_y).|\widehat{\phi}(\sigma.\omega_x,\sigma.\omega_y)|^2$$

et inversement,

$$c_\sigma = I * \frac{1}{\sigma^2}(\phi_\sigma^* * \phi_\sigma)$$

Finalement, si l'on somme la contribution de chaque échelle,

$$\begin{cases} I_{\texttt{rec}} = I * K \\ \text{avec } K = \sum_\sigma \frac{1}{\sigma^2}(\phi_\sigma^* * \phi_\sigma) \end{cases} \tag{3.2.7}$$

Cette relation est analogue à la transformation d'un point dans un système optique (*Point Spread Function*) et K va donc correspondre à la réponse impulsionnelle du filtre linéaire constitué par le codage / décodage à travers la transformée en ondelettes (et $\hat{K}$ symétriquement à la fonction de transfert de la transformée linéaire) . Il y a donc une coopération fine entre le sous-échantillonnage et la taille des filtres qui fait en sorte que la reconstruction soit simple. Aussi, ce théorème montre que la transformée ne doit pas être strictement une transformée en ondelettes, dans le sens où les filtres se déduisent par des dilations, et on choisira en particulier des filtres différents pour chaque taille afin de satisfaire des contraintes computationnelles.

Ce théorème montre aussi que la reconstruction ne sera parfaite que sous une condition sur les filtres, c'est-à-dire que le noyau K soit un Dirac discret (*i.e.* $K = 0$ sauf à l'origine $K(0,0) = 1$). Finalement, si et seulement si

$$\boxed{K := \sum_\sigma \frac{1}{\sigma^2}(\phi_\sigma^* * \phi_\sigma) = \delta_0} \tag{3.2.8}$$

Alors, la reconstruction sera parfaite[§]. Et seulement dans ce cas, on pourra écrire comme dans [Van Rullen et Thorpe, 2001] :

$$\boxed{I^T_{\texttt{rec}}(i,j) = \sum\nolimits_{1\le t\le T} < I, \phi_{i_t} > .\phi_{i_t}(\|(x_t - i, y_t - j)\|) \overset{T\to\infty}{\to} I} \tag{3.2.9}$$

Résultats de la reconstruction théorique de l'image

Pour garantir une bonne reconstruction de l'image par les coefficients de la transformée, on va donc ajuster l'ondelette "mère" qui permet de générer les filtres à toutes les échelles pour que K vérifie au mieux la condition de reconstruction. En pratique, il est nécessaire d'ajuster les filtres aux différentes

[§]On pourra écrire cette condition dans l'espace de Fourier : $\forall(\omega_x, \omega_y), \hat{K}(\omega_x, \omega_y) = 1$

tailles (la démonstration ci-dessus montre qu'on peut rompre la condition de la transformée en ondelettes et choisir des filtres différents à chaque échelle) pour qu'ils ne répondent pas à une surface plane, *i.e.* qu'ils gardent une somme nulle (en particulier aux petites tailles). Aussi on adaptera les filtres, notamment aux grandes tailles, pour qu'ils préservent une invariance par rotation. Répétons que même s'il est biologiquement exclu que l'image soit reconstruite dans le système visuel, nous allons étudier la qualité de cette reconstruction de l'image à partir de cette information pour mesurer la qualité de la transmission à travers le canal du nerf optique.

Aussi, cette architecture possède certaines limites. Tout d'abord, en rapport avec le traitement de l'image, des phénomènes au bord de l'image vont gêner une propagation homogène de l'image. En effet, si en théorie des ondelettes ceux-ci peuvent être pris en compte, les bords de l'image dans cette description vont correspondre aux contrastes les plus élevés et donc à des valeurs considérées comme les plus importantes. Les solutions classiques sont d'utiliser des convolutions circulaires, ou préférentiellement d'étendre l'image par miroir sur ses bords (ce que nous avons utilisé dans cette section pour pouvoir répliquer les résultats de reconstruction). Ensuite, un problème plus important va être la définition des filtres sur la discrétisation imposée par le maillage carré de l'image et pour la taille définie dans l'article. Concrètement, on observe à l'échelle 1 que le filtre gaussien de variance 1 est mal approché par la formule Éq. 3.2.2 et donc que le filtre de contraste est inadéquat (de composante constante non nulle, par exemple). On utilisera donc des approximations de ces formules sur des supports compacts pour que la reconstruction soit la plus correcte.

Aussi pour les filtres de plus grande échelle, les calculs deviennent très onéreux. Une solution naturelle est alors de construire une *Pyramide Laplacienne* [Burt et Adelson, 1983]. Cet algorithme est une transformée multi-échelle linéaire définie sur une grille dyadique qui permet de calculer les coefficients d'une transformée de type ondelettes à différentes échelles par récurrence d'une échelle à l'autre à partir de filtres élémentaires 3×3 ou 5×5. Aussi, par son architecture, le sous-échantillonnage puis le sur-échantillonnage sont ajustés pour que la reconstruction depuis les coefficients soit parfaite : on obtient

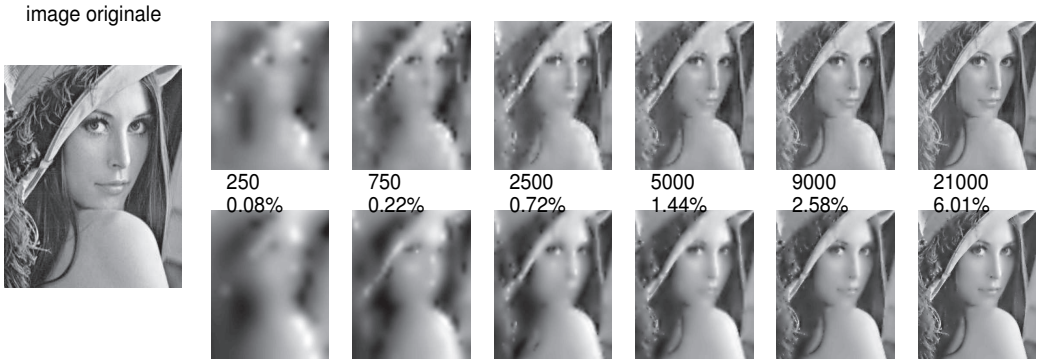

FIG. 3.2.3 – *Reconstruction progressive de l'image par les coefficients de la transformée.* Pour étudier la qualité du code impulsionnel, l'image est reconstruite théoriquement au niveau du CGL en supposant connue la valeur des coefficients pour différentes transformées. Résultat sur un détail de l'image "Lena" pour **(Haut)** la transformée en ondelettes définie par [Van Rullen et Thorpe, 2001] et **(Bas)** la pyramide laplacienne. L'information est transmise progressivement par les impulsions (resp. de gauche à droite après 250, 750, 2500, 5000, 9000 et 21000 impulsions) de telle sorte qu'après 1% de neurones activés, on reconnait nettement l'image. Les deux transformées semblent de performances équivalentes, toutefois la complexité computationnelle de la pyramide laplacienne est largement moindre.

---

un codage de l'image en *sous-bandes* qui vont correspondre au découplage des différentes échelles dyadiques de grandeur des objets. Nous avons mis en œuvre cet algorithme grâce aux travaux de Eero Simoncelli (et décrits dans [Simoncelli et Freeman, 1995]) ; la charge computationnelle et l'utilisation de la mémoire est alors largement réduite en comparaison avec la transformée en ondelettes classique. Enfin, cet algorithme est biologiquement plausible car ce calcul récursif de filtres peut être mis en correspondance avec les processus de diffusion liés aux processus chimiques et aux jonctions synaptiques électriques dans la rétine : elle introduit ainsi un principe d'économie du volume de connexions et montre que les interactions locales vont pouvoir implanter de proche en proche le calcul de contrastes même pour de grandes échelles.

En pratique, on a dans un premier temps répliqué l'algorithme utilisant les

ondelettes, puis l'algorithme équivalent de propagation utilisant les coefficients de la pyramide laplacienne. Dans les deux cas, on reconnaît l'image après seulement quelques centaines de PAs (sur un total de $512^2 * 8/3 \sim 700.10^3$, voir Fig. 3.2.3) soit dans un cadre physiologique environ 25 ms, et la reconstruction est presque parfaite après qu'autant de PAs aient été émis qu'il y a de pixels (soit environ 20%). On remarque la pyramide laplacienne semble aussi adaptée, ce qui est confirmé par les mesures statistiques de la convergence de ces reconstructions sur une base d'images[¶] dont on extrait des imagettes de 128×128 pixels. On mesure alors en général la convergence de l'algorithme en mesurent l'Erreur Quadratique Moyenne (*EQM* ou en anglais Mean-Squared Error, *MSE*) qui mesure quantitativement une énergie moyenne de l'écart entre l'image I et sa reconstruction $I_{\texttt{rec}}$ sur les pixels $\mathcal{I}$ de l'image :

$$EQM(I, I_{\texttt{rec}}) = E[|I - I_{\texttt{rec}}|^2] = E[\sum\nolimits_{\vec{l} \in \mathcal{I}} (I(\vec{l}) - I_{\texttt{rec}}(\vec{l}))^2] \qquad (3.2.10)$$

et l'Information Mutuelle, comme la quantité moyenne d'information gagnée sur I quand $I_{\texttt{rec}}$ est connu soit la somme des entropies (voir Éq. 3.1.28) des luminosités pour chaque image L_I et $L_{I_{rec}}$ (donc l'entropie des probabilités marginales) moins l'entropie des luminosités couplées (soit l'entropie de la probabilité jointe) :

$$\begin{aligned} IM(I, I_{\texttt{rec}}) &= H[P(I)] + H[P(I_{\texttt{rec}})] - H[P(I, I_{\texttt{rec}})] & (3.2.11) \\ &= \sum_{\substack{L_I \in \mathcal{L}_{\mathcal{I}} \\ L_{I_{\texttt{rec}}} \in \mathcal{L}_{I_{\texttt{rec}}}}} P(L_I, L_{I_{\texttt{rec}}}).log(\frac{P(L_I, L_{I_{\texttt{rec}}})}{P(L_I).P(L_{I_{\texttt{rec}}})}) & (3.2.12) \end{aligned}$$

L'information mutuelle sera donc nulle pour des valeurs indépendantes et sera maximale (égale à l'entropie de la luminosité dans l'image) pour une reconstruction exacte. Ces mesures donnent des résultats relativement similaires pour les deux méthodes (voir Fig. 3.2.4), montrant ainsi qu'une reconstruction progressive par les coefficients conduit bien à une reconstruction progressive de l'image. Mais les résultats, comparés à la charge computationnelle de

[¶]Cette base, utilisée par Hoyer et Hyvãrinen [2002], est rendue disponible par l'auteur. Nous utiliserons ce même protocole pour comparer les modèles proposés.

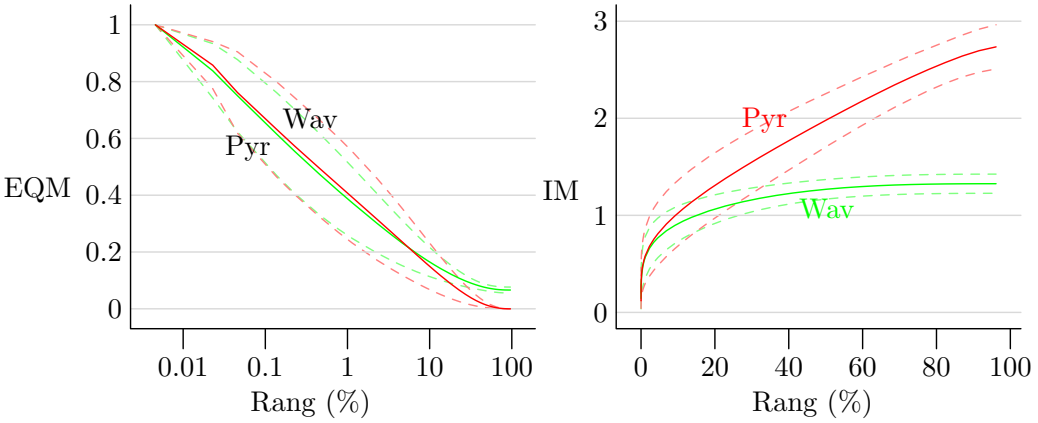

FIG. 3.2.4 – *Mesure de la convergence de la reconstruction progressive de l'image par les coefficients de la transformée.* On a comparé pour les deux transformées (ondelettes en trait vert et pyramide laplacienne en rouge) la mesure de **(A)** l'Erreur Quadratique Moyenne (EQM ou MSE) et **(B)** de l'Information Mutuelle (l'écart type de ces mesures est indiqué par les traits pointillés). La reconstruction finale est exacte pour le pyramide laplacienne, et la reconstruction progressive est équivalente au sens de l'EQM mais la pyramide laplacienne est supérieure pour l'information mutuelle car cette dernière mesure est particulièrement sensible aux erreurs de flou visuel.

chaque méthode, suggèrent que la pyramide laplacienne permet de proposer à la fois un modèle plus simple et plus rapide mais aussi de performance supérieure.

3.2.2 Résultats : un code impulsionnel compact

Transformation en un train de PAs

En suivant le modèle de [Van Rullen et Thorpe, 2001], les activités que nous avons calculées vont correspondre à un courant excitateur appliqué aux cellules ganglionnaires, et ces cellules vont donc produire des impulsions d'autant plus rapidement qu'elles sont activées fortement. Les sous-couches ON et OFF vont donc fournir un signal impulsionnel qui va coder les images des coefficients de la transformée rétinienne. Comme évoqué dans Sec. 3.1.3, nous allons nous concentrer sur la première impulsion émise par chaque neurone

et l'image va donc être transformée en un *front d'impulsions* le long du nerf optique et nous allons alors tout d'abord étudier la décomposition de l'information grâce aux PAs générés par les cellules ganglionnaires.
Si l'on observe alors le signal suivant le décours temporel et comme la transformation de l'activité en latence est monotone, alors on observera les coefficients de l'image "flashée" depuis le plus actif jusqu'au moins actif. On pourra alors plutôt considérer le signal comme une liste de PAs, et au lieu de considérer l'activité de chaque neurone, on observera alors plutôt la succession dans le temps des fibres activées. On comprend alors qu'il est important de fixer précisément la norme des poids synaptiques des différents filtres (c'est-à-dire le gain des cellules ganglionnaires) car elle contrôlera la latence relative des filtres. Dans un premier temps, comme dans [Van Rullen et Thorpe, 2001] et comme il est habituel en théorie des ondelettes, nous choisissons de normaliser les filtres pour n'importe quelle taille à 1, c'est-à-dire de poser que si nous présentons "l'image" des poids d'un filtre quelconque, la réponse de ce filtre à lui même sera 1. L'image d'un filtre quelconque (quel que soit sa taille) fournira un premier PA à une même latence standard.
Nous considérerons alors que le nerf optique n'introduit pas de différence de délai et que la volée de PAs produite à la rétine arrive dans le même ordre au CGL. Une fois ce schéma défini, on peut se demander comment les impulsions transmises au CGL peuvent être interprétées et comment elles peuvent transmettre la valeur des coefficients de l'ondelette et donc une reconstruction progressive de l'image. On peut alors proposer que le *rang* de ces PAs est suffisant pour reconstruire correctement l'image. Pour cela, à l'image d'une analyse de corrélation inversée, nous allons construire une fonction donnant la valeur analogique en fonction du rang et évaluer le code transmis. S'il est validé, ce code va donc permettre de transmettre l'information visuelle par une transformation impulsionnelle.

Transmission de la valeur des coefficients par leur ordre

En fait, si l'on note $x(\frac{k}{N})$ la valeur absolue du coefficient de rang k (pour N coefficients) et $z = \frac{k}{N}$ donc la proportion de coefficients supérieurs en norme

à $x(z)$, alors pour les images naturelles qui sont régulières par morceaux, on observe (voir [Mallat, 1998, p.513] qu'après un certain rang,

$$\gamma(z) = -\frac{d \log_2 x(z)}{d \log_2 z} \sim 1 \tag{3.2.13}$$

On pourra donc écrire (pour les coefficients correspondant à des parties régulières) $x(z) \sim C.z^{-\gamma}$ avec $\gamma \sim 1$. En fait, cette relation pourra se généraliser car le taux de décroissance de x est directement liée aux coefficients de Lipschitz (les transformations par ondelettes donnant une évaluation locale de ce coefficient dans l'image [Mallat et Hwang, 1991]) dans l'image— c'est-à-dire à des mesures de l'irrégularité de façon décroissante : impulsion, créneau , pente, plateau— et donc que cette fonction va évaluer la moyenne de la répartition des discontinuités des images naturelles. Cette relation vérifie théoriquement les observations précédentes. Elles soulignent aussi le fait que les plus larges coefficients ($z << 1$) sont créés par des discontinuités, pour lesquelles $\gamma(z) > 1$ et que celles-ci seront propagées par ordre décroisant de discontinuité. Les premières valeurs transmises correspondent alors à la fois à une grande importance (ils transmettent les contours) et à une grande variabilité.

En effet, Van Rullen et Thorpe [2001] a observé que la fonction donnant la valeur absolue des coefficients en fonction du rang que l'on a utilisé pour définir l'ordre de propagation des PAs était régulière sur une base de donnée d'images naturelles. On peut donc construire une fonction donnant la *modulation* en fonction du rang en calculant la moyenne de la valeur absolue des coefficients —ensuite normalisée à 1 au rang 1— pour déduire la valeur de coefficient d'un PA depuis son rang d'arrivée. Cette normalisation induit donc que la reconstruction sera approchée à un coefficient multiplicateur près. La méthode pour retrouver ce coefficient n'est pas précisée dans l'article original.

On peut donc explorer différentes stratégies pour décoder la valeur absolue d'un coefficient en fonction de son rang. Tout d'abord on peut prendre la moyenne de ces fonctions pour une base de donnée d'images naturelles (voir Éq. 3.2.14 où l'on a noté $\langle y \rangle$ la moyenne de la variable y sur cette base), et

on obtient une fonction comme une table arithmétique (*Look-Up-Table* ou LUT) donnant la décroissance de la valeur absolue du contraste en fonction du rang (LUT_1). On peut aussi imaginer calculer cette LUT en normalisant les fonctions à 1 au rang 1 (*i.e.* pour $z = 0$, LUT_2) puisqu'elle donne plus de précision aux premiers coefficients, *i.e.* les plus importants (la variance sera alors nulle au rang 1). On pourra enfin comparer ces deux méthodes à une méthode normalisant chaque fonction en fonction de l'énergie de l'image, c'est-à-dire à la déviation standard des coefficients (LUT_3), cette dernière méthode se justifiant biologiquement par l'existence d'un tel calcul de variance par les cellules ganglionnaires de type "Y".

$$\begin{cases} LUT_1(z) = \langle x(z) \rangle \\ LUT_2(z) = \langle x(z)/x(0) \rangle \\ LUT_3(z) = \left\langle x(z)/\sqrt{\sum x^2} \right\rangle \end{cases} \tag{3.2.14}$$

On observe alors que ces méthodes donnent des courbes sensiblement semblables (voir les lignes de Fig. 3.2.5), mais que la variance peut différer nettement (les zones pleines de la figure correspondent à l'écart depuis la moyenne d'une fois l'écart type). On préférera en général une variance faible, mais comme l'importance des coefficients varie avec le rang, il faudra évaluer ces stratégies par des simulations.

Aussi, comme cette règle est relativement stable sur les images de la base de donnée, nous avons aussi choisi de construire une règle plus biologique. Celle-ci est basée sur un algorithme d'apprentissage stochastique : donnée une initialisation aléatoire $m^{(0)}$ de la modulation* après le n$^{\text{ème}}$ codage utilisant $m^{(n)}$ comme fonction de modulation (*i.e.* en estimant $|C^t_{i^t}|$ par $m^{(n)}(t)$,

$$m^{(n+1)}(t) = (1 - \mu^{(n)}).m^{(n)}(t) + \mu^{(n)}.|C^t_{i^t}| \tag{3.2.15}$$

avec $\mu^{(n)}$ le gain stochastique† et qui vérifie donc $\sum_{n=1\ldots\infty} \mu_n \rightarrow \infty$ et $\sum_{n=1\ldots\infty} {\mu_n}^2 < \infty$. Cette règle peut être réalisée biologiquement par un pro-

*toutefois décroissante, on choisit e.g. la règle linéaire avec $m^{(0)}(1) = 1$ et $m^{(0)}(N) = 0$

†typiquement, $\mu^{(n)} = 1/n$

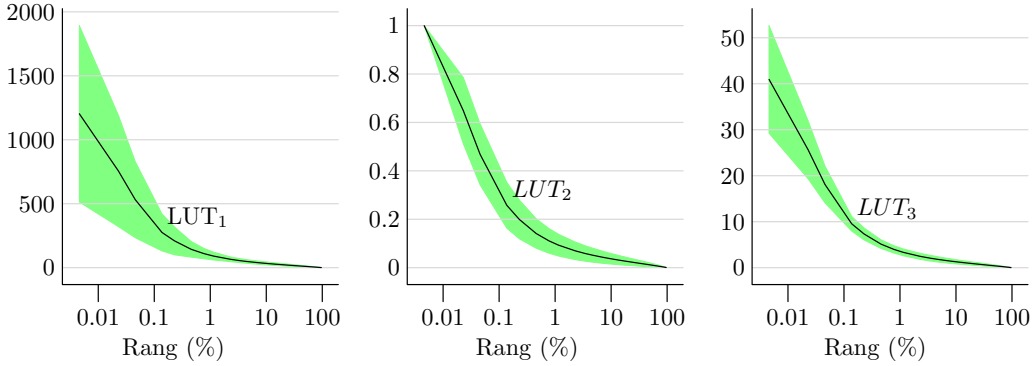

FIG. 3.2.5 – *Quantification par leur rang de la valeur absolue des coefficients de la transformée.* On a comparé les trois fonctions de quantification **(A)** LUT_1, **(B)** LUT_2 et **(C)** LUT_3 (voir Éq. 3.2.14) les zones pleines correspondant à un écart d'une fois l'écart type depuis la moyenne. Les dernières méthodes réduisent nettement la variance mais nécessite la transmission d'un coefficient normalisateur.

cessus adaptatif à plus long terme que la dynamique du système et peut par exemple utiliser le codage par fréquence. Comme mathématiquement, le choix de μ modifie seulement le décodage et non le codage, μ tend donc vers la fonction LUT_1 calculée précédemment, ce qui est confirmé par les simulations.

Une représentation compacte et dynamique

La reconstruction d'imagettes 128×128 par cette méthode montre que la reconstruction de l'image par la valeur quantifiée des coefficients dans une pyramide laplacienne est relativement bonne (voir Fig. 3.2.6). Les différentes stratégies donnent des résultats équivalents à ce stade de l'étude et montrent que ce code peut être utilisé effectivement depuis la rétine pour coder les valeurs analogiques à partir de leur rang de tir. En effet, nous observons un résultat très important pour notre modèle de la CUR : le coefficient correspondant à un PA qui est nécessaire à la reconstruction de l'image dans le CGL peut être donné avec une bonne précision par le rang de ce PA et la reconstruction correspondante est proche de la reconstruction progressive idéale.

Ce modèle permet donc d'envisager le codage dynamique de la transmission d'une image en un front d'impulsions et qui soit complètement compatible avec les restrictions imposées par la CUR.
Finalement, on a donc implanté un code visuel impulsionnel complet qui permet de coder une image par une volée d'impulsions et de la décoder simplement par leur ordre de décharge. Les améliorations et extensions que nous présentons montrent que ce modèle peut s'adapter à différentes architectures, et particulièrement à des algorithmes plus rapides et plausibles. En particulier, la transformation par une pyramide laplacienne va permettre de proposer une architecture simple et plus performante. En particulier, la théorie des ondelettes montre que ces transformées sur des bases de filtres orthogonaux (ou "presque") vont permettre d'implanter un code *dense* de l'image. La propagation progressive des coefficients généralise alors les notions de réduction de dimension, de compression et de quantification à des processus dynamiques qui sont caractéristiques de la rétine.
L'utilisation d'un codage dense dans ce modèle du code rétinien présente des limites. D'un côté, les filtres rétiniens biologiques ne forment pas une base orthogonale [Salinas et Sejnowski, 2001]. Aussi, d'après la théorie des ondelettes et même avec cette base non-orthogonale, le code obtenu sera au mieux compact alors que des mesures neurophysiologiques suggèrent que la décroissance des coefficients est plus rapide. Nous étudierons de tels modèles au Chap. 4 en implantant des architectures similaires mais avec des interactions latérales. Toutefois, ce modèle va nous permettre de définir un code impulsionnel complet et d'étudier son comportement dynamique.

3.2.3 Optimisation statistique du modèle dynamique rétinien par les statistiques des images naturelles

En pratique, on observe que la propagation de l'information visuelle dans cette architecture laisse apparaître des régularités qui suggèrent l'importance d'une optimisation statistique des paramètres du modèle. En effet, l'implantation de ce modèle de code rétinien montre des comportements semblables pour des images différentes et de manière générale, les basses fréquences vont

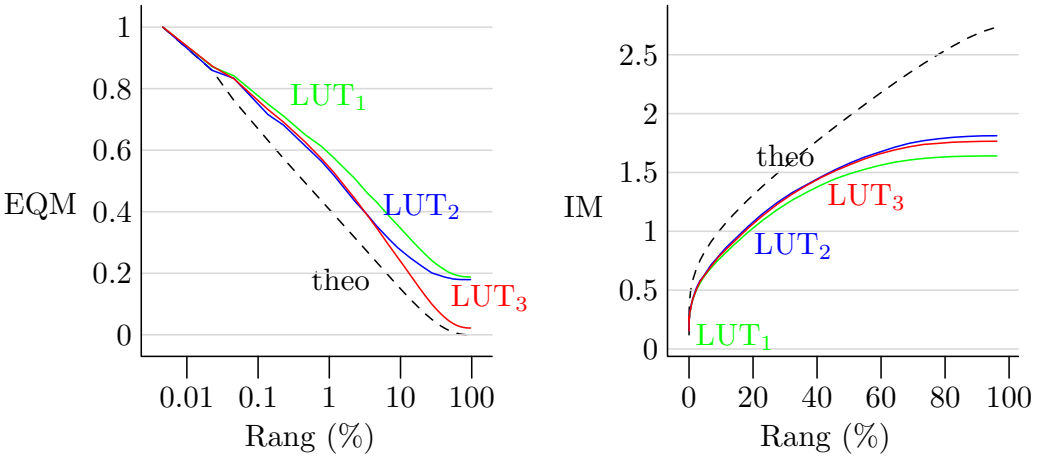

FIG. 3.2.6 – *Mesure de la convergence de la reconstruction progressive de l'image par les coefficients de la transformée.* On a comparé pour les différentes quantification des coefficients la mesure de **(A)** l'Erreur Quadratique moyenne (EQM ou MSE) et **(B)** de l'information mutuelle relativement à la reconstruction progressive théorique (trait pointillé, Fig. 3.2.4. Les différentes stratégies sont relativement équivalentes même si LUT_3 donne une meilleure reconstruction finale.

plus souvent correspondre à de hautes valeurs et seront propagées en premier.

Décorrélation spatiale

Comme nous avons étudié les régularités de la luminance pour définir la luminosité, on peut s'intéresser à la dépendance de la luminosité de deux points $\vec{x}$ et $\vec{y}$ en fonction de leur position. Ces études sont à la base de l'étude des *statistiques des images naturelles* et cherchent à trouver des invariances dans les images pour en déduire des stratégies optimales de codage. Une première mesure statistique mesurée est alors l'auto-corrélation spatiale bidimensionnelle des luminosités (préalablement centrées autour de la moyenne) :

$$R(\vec{x},\vec{y}) = \langle L(\vec{x}).L(\vec{y})\rangle - \langle L(\vec{x})\rangle \,.\, \langle L(\vec{y})\rangle \tag{3.2.16}$$

qui est la moyenne sur de nombreuses images de la corrélation des luminosité L en ces points. Par homogénéité par translation dans les images naturelles,

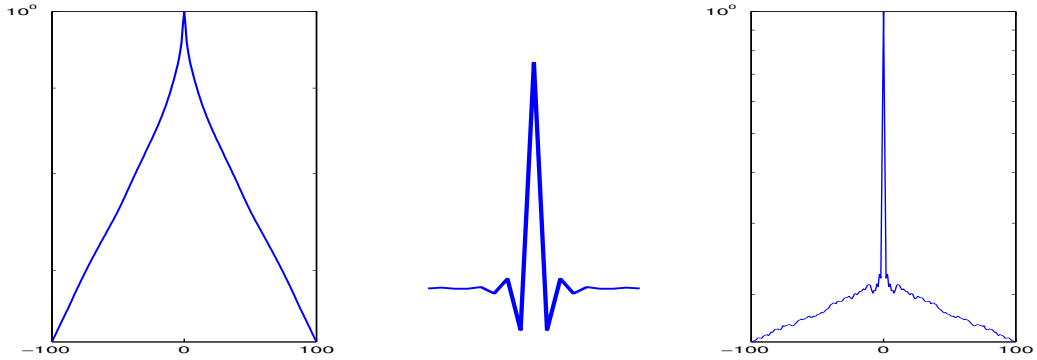

FIG. 3.2.7 – *Décorrélation spatiale.* **(Gauche)** Spectre spatial moyen d'énergie. Celui-ci est calculé sur une base de donnée d'images naturelles : le spectre d'énergie décroît en $\frac{1}{f^2}$. **(Milieu)** Filtre de décorrélation K calculé d'après les méthodes de Atick [1992] et Olshausen et Field [1998]. Ce filtre se rapproche de la transformation dans la première couche de la rétine par les cellules bipolaires et horizontales. **(Droite)** Spectre spatial moyen d'énergie après filtrage : les luminosités des points en fonction de leur distance sont donc décorrélées par K.

---

on note $R(\vec{x}-\vec{y}) = R(\vec{x},\vec{y})$ et on en déduit par la transformée de Fourier $\hat{R}$ de R le *spectre spatial de puissance** : $\hat{R}(\vec{f}) = \int d\vec{x} e^{i\vec{f}\vec{x}} R(\vec{x})$. En comparant ce calcul sur différentes images, Field [1987] a alors trouvé que $\hat{R}$ obéissait à

$$\hat{R}(\vec{f}) \sim \frac{1}{\left\|\vec{f}\right\|^2} \tag{3.2.17}$$

ce qui correspond à un auto-corrélateur invariant par changement d'échelle (*i.e.* $\hat{R}(\vec{f}) = \alpha^2.\hat{R}(\alpha.\vec{f})$. Cette propriété découle du fait qu'a priori, les objets peuvent être à des tailles aléatoires, mais qu'alors leur nombre varie inversement proportionnellement au carré de leur taille [Alvarez et al., 1999].

Une stratégie de codage pour la rétine introduite par Atick [1992] et Atick et Redlich [1992] est donc pour les cellules ganglionnaires de fournir une réponse décorrélée et est donc reliée à une stratégie de maximisation de l'entropie de l'activité des cellules ganglionnaires. Pour cela, en considérant cette activité comme la transformée linéaire $O =< L, K >$, avec L l'image des luminosités

*on note que $\hat{R}$ est symétrique positive

et K le filtre rétinien, on peut déduire le spectre de puissance de O comme le produit de ceux de L et K. On en déduit que si l'on veut que la sortie soit décorrélée, il suffit de poser

$$\left\|\hat{K}(\vec{f})\right\| \propto \sqrt{\frac{1}{\left\|\hat{R}(\vec{f})\right\|}} \tag{3.2.18}$$

soit

$$\left\|\hat{K}(\vec{f})\right\| \sim \left\|\vec{f}\right\| \tag{3.2.19}$$

Toutefois, cette stratégie va considérablement augmenter les hautes fréquences et donc le bruit lié aux récepteurs. Il est donc nécessaire "d'adoucir" les images (soit donc K par associativité de l'opération de convolution) un filtre spatial passe-bas B. D'après des données psychophysiques, on choisit $\hat{B} = \exp(-(\frac{\|f\|}{\|f_c\|})^{\alpha})$, où $f_c \sim 22\text{deg}^{-1}$ et $\alpha \sim 1.4$ [Olshausen et Field, 1998]. Finalement, on obtient un filtre rétinien $K \leftarrow K * B$ (voir Fig. 3.2.7-Milieu).
Leur étude peut s'étendre à notre modèle multiéchelle car comme le système est linéaire jusqu'à la génération des PAs (par les cellules ganglionnaires), on pourra composer ce filtre de décorrélation avec nos filtres aux différentes échelles pour retrouver une architecture similaire à notre modèle. De plus, comme les filtres d'ondelettes sont relativement localisés autour de leur fréquence spatiale préférée, on peut utiliser en première approximation la norme de la convolution de ce filtre avec K pour mesurer la modification de la sensibilité des filtres avec leur taille. Les résultats montrent alors des résultats similaires pour ces deux méthodes et on peut noter que les normes relatives des filtres rétiniens de ce modèle vont correspondre aux mesures physiologiques d'Enroth-Cugell et Robson [1966]. En pratique, on observe alors que sur une base d'images, par rapport à la situation précédente où les coefficients de filtres de grande taille étaient a priori plus forts, les histogrammes des coefficients aux différentes échelles sont harmonisés (voir Fig. 3.2.8).

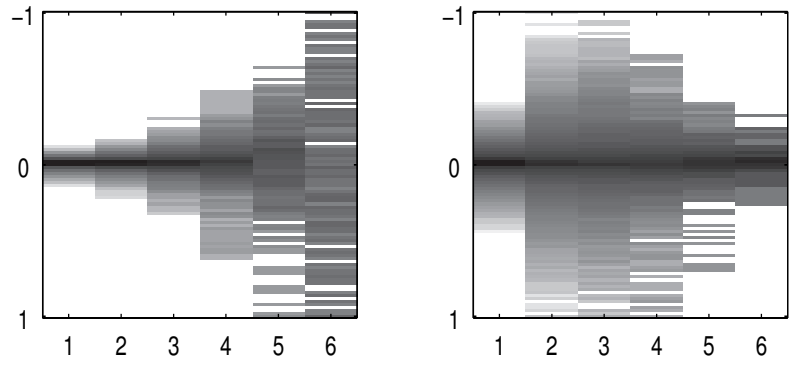

FIG. 3.2.8 – *Harmonisation des histogrammes par la statistique des images naturelles.* Logarithme de l'histogramme moyen des coefficients d'une représentation multi-échelles dyadique (pyramide laplacienne) en fonction de l'échelle (6 correspondant à une taille de 2^6, une couleur sombre correspondant à une probabilité forte, soit une information faible) $(\mathbf{A})$ sans ou $(\mathbf{B})$ avec décorrélation. Les énergies sont mieux réparties et vont donc correspondre à une plus grande information par coefficient. Le bruit, concentré sur les premières échelles est atténué par un filtre passe-bas. Dans le modèle progressif de propagation, l'ordre des filtres choisis en les rangeant par activité seront alors indépendant de la taille.

Distance de Mahalanobis

Nous avons vu que nous pouvons modifier l'importance relative des différentes échelles pour maximiser l'entropie du code rétinien en reliant la sensibilité de le norme des neurones inversement proportionnelle à leur occurrence. Cette stratégie se rapproche de la distance introduite par Mahalanobis [1936] qui permet de définir une distance pour des vecteurs à plusieurs variables n'ayant souvent pas de rapport direct entre-elles, comme c'est souvent le cas en statistique. Elle est définie alors pour deux vecteurs $\vec{x}$ et $\vec{y}$ par

$$\mathcal{D}_M(\vec{x}-\vec{y}) = (\vec{x}-\vec{y})\Sigma^{-1}(\vec{x}-\vec{y})^T \tag{3.2.20}$$

où Σ correspond à la matrice de covariance de la variable $\vec{x}$. Dans notre cas, cette méthode se rapproche de notre calcul du filtre de décorrélation K (voir Éq. 3.2.18), et correspond à la même stratégie de renormalisation de chaque composante de l'information (pour l'image, cela correspondra aux différentes échelles) pour pouvoir les comparer et donc définir une distance. On peut aussi rapprocher cette stratégie de l'analyse en composantes principales (ACP), puisque nous avons décorrélé la valeur de la luminosité en fonction de la taille en normalisant chaque composante (ici les tailles) pour que la densité de probabilité moyenne sur les images naturelles soit étalée (soit sphérique si l'on considère l'espace multidimensionnel des luminosités pour des axes correspondant aux différentes tailles).
Nous avons donc défini une distance sur les images qui rééquilibre les différentes composantes d'échelle. Elle peut se coupler avec l'Erreur Quadratique Moyenne (*Mean-Squared Error*, MSE) pour définir une nouvelle distance entre deux images comme l'Erreur Quadratique Moyenne Balancée (EQMB, ou *Weighted Mean-Squared Error*, WMSE). Si l'on considère le filtre de blanchiment K, on peut calculer de façon équivalente l'EQMB comme l'EQM des images blanchies ou comme la valeur quadratique moyenne de l'image blanchie de l'erreur.

$$\begin{aligned} EQMB(I, I_{\texttt{rec}}) &= E[|(I - I_{\texttt{rec}}) * K|^2] && (3.2.21)\\ &= E[|I * K - I_{\texttt{rec}} * K|^2] && (3.2.22) \end{aligned}$$

Cette distance donne une évaluation quantitative qui correspond mieux à une distance subjective entre deux images. Pour preuve, la figure Fig. 3.2.9 montre une image et 5 versions de celle-ci bruitées selon des fréquences différentes. On a alors évalué les distances normalisées suivant les deux méthodes présentées. La EQMB correspond alors nettement plus au rang subjectif que celui donné par la EQM. Ces résultats montrent qu'un bruit dont la puissance spectrale en fréquence diminue en $1/f$ est celui qui perturbe le moins la perception pour une puissance de bruit donnée par rapport à un bruit blanc, et que la distance EQMB est donc parfaitement adaptée à la mesure de la distance entre des images naturelles et donc à l'évaluation du codage rétinien. De façon annexe, comme évoqué dans Van Rullen et Thorpe [2001], alors que la MSE est sensible à la composante continue de l'image, ce n'est plus le cas pour la EQMB[†].

Code impulsionnel optimisé statistiquement

Nous avons alors implanté ces résultats pour l'architecture de code rétinien impulsionnel en changeant simplement la norme des filtres aux différentes échelles. On observe alors que l'on ne peut pas prévoir quelle échelle va être propagée en premier lieu et donc que les filtres peuvent correspondre à de petites tailles dès les premiers rangs. En particulier, la reconstruction de l'image va faire apparaître les contours (et les discontinuités) en premier lieu. En fait, on s'aperçoit de façon générale que les zones de l'image vont être maintenant propagées suivant des irrégularités décroissantes et indépendamment de leur taille. Ce codage rétinien permet donc de propager en avant et dans l'ordre les points puis les lignes, créneaux et pentes qui composent l'image.

Il est alors intéressant de suivre la même procédure que plus haut afin de construire un code impulsionnel complet. En effet, en calculant de la même façon la valeur absolue du coefficient en fonction du rang, on s'aperçoit maintenant que même pour LUT_1, la variance est largement réduite (voir Fig. 3.2.10) et que le graphe de la modulation obéit à une loi linéaire du logarithme du rang relatif. En comparant ces courbes aux différentes échelles,

[†]En effet, l'auto-corrélation de la luminosité des points dans une image est infinie à l'origine et la pondération proposée annule donc cette composante

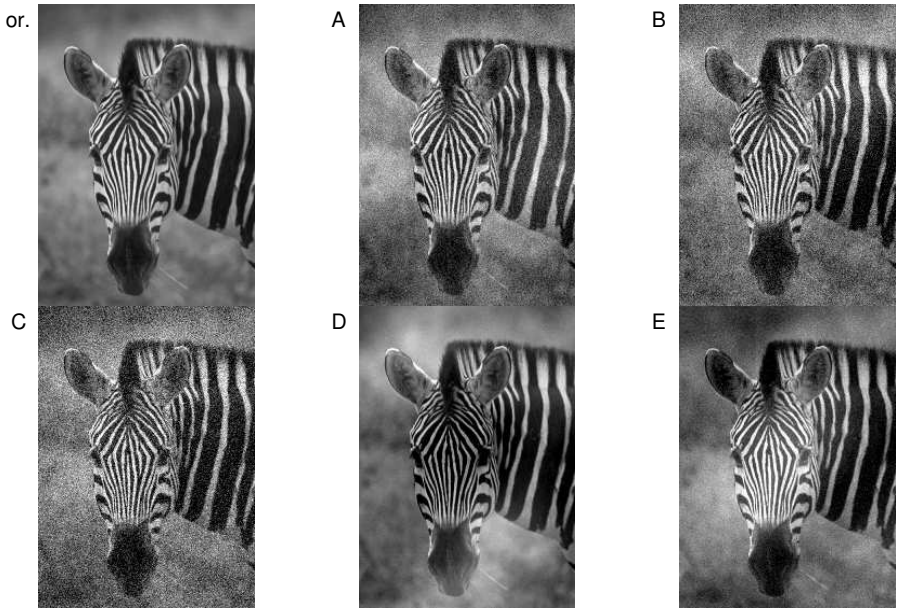

FIG. 3.2.9 – *Distance subjective entre des images.* On a bruité une image originale **(Or.)** par des bruits de spectres différents pour obtenir 5 images différentes **(A)** à **(E)**. Les mesures de distance donnent alors la distance depuis l'original : pour la MSE les résultats sont du plus proche au plus éloigné A, B, E, C et D alors que EQMB donne D, E, A, B et C. Cette dernière distance correspond donc plus à une mesure subjective du bruit ajouté sur l'image originale (voir par exemple en détail l'oreille du zèbre) et par extension de la distance entre des images.

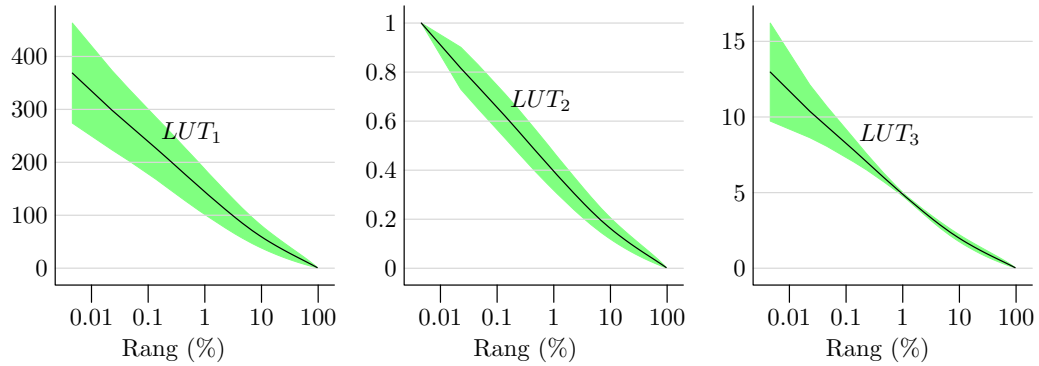

FIG. 3.2.10 – *Quantification de la valeur absolue des coefficients de la transformée par leur rang.* On a comparé les trois fonctions de quantification **(A)** LUT_1, **(B)** LUT_2 et **(C)** LUT_3 (voir Éq. 3.2.14) avec la déviation standard lors de l'apprentissage (la zone pleine correspond à une déviation standard de la moyenne). Les variances sont plus faibles que dans Fig. 3.2.5, et les courbes sont rectilignes, suggérant une loi en puissance de la répartition des discontinuités dans les images naturelles.

---

on observe que dans le cas précédent, les décroissances des coefficients sont différentes, alors que grâce à l'optimisation statistique, on a ramené cette décroissance des coefficients aux différentes tailles à des amplitudes et pentes similaires. Ceci explique donc que l'utilisation de la LUT est ici plus naturel dans ce cas car il correspond à une régularité dans la physique des images naturelles ; cette loi reflète la distribution de probabilité des coefficients d'irrégularité (de Lipschitz) dans les images naturelles et est liée aux propriétés d'auto-similarité des images naturelles [Turiel et al., 1998; Turiel et Parga, 2000].

On en déduit de ces calculs une implantation qui permettent alors de déterminer les résultats de la propagation progressive de l'image pour ce code optimisé (voir Fig. 3.2.11). Comme le laissait prévoir nos observations et la faible variance de la fonction de modulation, on observe que la convergence de la reconstruction est beaucoup plus fine que pour la reconstruction précédente. Aussi, on remarque que si on découple la propagation à différentes échelles (ce qui semble être le cas dans la rétine avec les voies Magno et Parvo), on pourra garder cette stratégie de codage de la valeur

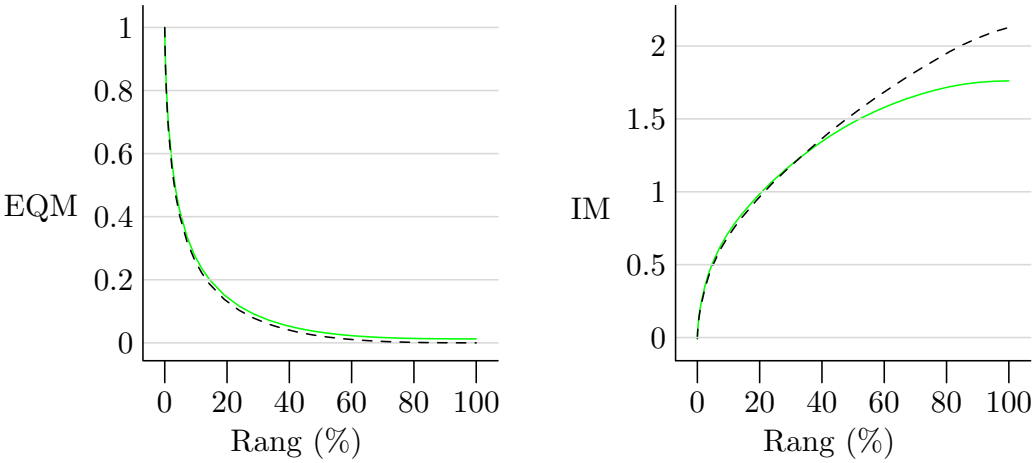

FIG. 3.2.11 – *Mesure de la convergence de la reconstruction progressive de l'image par les coefficients de la transformée optimisée.* On a comparé la mesure de **(A)** l'Erreur Quadratique moyenne (EQM ou MSE) et **(B)** de l'information mutuelle pour le code rétinien optimisé avec la reconstruction idéale par les coefficients. La reconstruction est plus lente au sens de l'EQM (car cette propagation se base sur l'EQMB) mais est beaucoup plus fine par rapport à la reconstruction théorique (traits pointillés). L'information mutuelle s'approche ainsi de façon quasi-parfaite de la reconstruction théorique.

absolue tout en propageant certaines échelles avant suivant les contraintes de rapidité imposées au système. Cette stratégie montre en effet que la valeur analogique est parfaitement codée par l'ordre d'activation des neurones dans cette architecture correctement équilibrée grâce aux statistiques des images naturelles.

3.3 Codage par Rang

Le modèle que nous avons construit présente donc un code complet de propagation asynchrone et temporelle de l'information rétinienne par l'ordre d'activation des neurones. Il permet grâce au *codage par rang* de déduire un mécanisme simple pour transmettre progressivement l'image au niveau du CGL. Cette alternative aux codages neuronaux classiques —et particulièrement au codage par la fréquence de décharge— a été proposée initialement par Thorpe et Imbert [1989] et permet dans un cadre biologiquement plausible d'expliquer le traitement rapide dans le système visuel. Mais c'est aussi un formidable outil de calcul parfaitement adapté au traitement du signal et de l'image.

3.3.1 Inspiration biologique du codage par rang

En se basant sur des données neurophysiologiques récentes, nous allons tout d'abord proposer un modèle qui complétera les modèles de la Sec. 3.1 en prenant en compte l'aspect temporel et asynchrone du signal neuronal.

Implantation biologique d'un codage par l'ordre d'activation

On a vu (voir Sec. 2.3) des mécanismes possibles pour qu'une partie de l'information impulsionnelle soit codée par la latence (voir Sec. 3.1) et nous avons montré que les paramètres des neurones et de leurs synapses étaient des facteurs déterminants pour prévoir leur réponse au flux impulsionnel. En particulier, les mécanismes d'adaptation que nous avons introduits induisent des mécanismes homéostatiques qui font en sorte que la réponse en latence est plus rapide (le potentiel étant plus souvent proche du seuil de décharge) et que

la réponse impulsionnelle moyenne des neurones reste très irrégulière. En fait, les rythmes d'activité neurale à l'échelle du SNC, comme le rythme γ, sont des facteurs qui soutiennent une propagation par "vagues" des impulsions neuronales. On définira alors grâce à ce rythme "l'initialisation" de la vague d'impulsions dans une assemblée neuronale et à partir de laquelle on déduira à chaque oscillation un vecteur des latences d'arrivée des impulsions sur un neurone. Comme nous l'avons vu pour notre modèle de la rétine, nous allons ici plus particulièrement étudier le vecteur des rangs d'arrivée des premières impulsions, soit le "front d'impulsions".

De plus, la section précédente a mis en évidence que des mécanismes écologiques qui modifient la sensibilité des neurones —en équilibrant par exemple la densité de probabilité de la latence de tir— peuvent être déterminants pour construire un code neural par le rang des impulsions comme celui qui a été mis en évidence dans la rétine (voir Sec. 3.2.2). Comme pour une commissure de N fibres, le nombre de signaux différents est du nombre de l'ensemble des permutations des fibres qui sont actives, soit de l'ordre de $e.N!$ où e est le nombre d'Euler*. Dans le cas où les probabilités sont parfaitement équilibrées, les vecteurs des rangs d'arrivée d'un front d'impulsions sont par conséquent uniformément répartis dans l'espace des permutations et alors dans ce cas, l'information portée par chaque signal du codage par rang est potentiellement de $\log_2(e.N!)$ bit.

Toutefois, on a vu que pour tenir compte du caractère asynchrone du signal, il est nécessaire de pouvoir évaluer le nombre d'impulsions qui sont arrivées de telle sorte qu'une décharge massive de toutes les fibres ne produise pas de réponse par rapport un signal préféré constitué d'un sous-ensemble de ces fibres. En fait, il semble que pour moduler l'efficacité synaptique suivant le rang, l'inhibition divisive (ou *shunting inhibition*, [Borg-Graham et al., 1998; Monier et al., 1997]) soit un mécanisme adapté. En effet, cette inhibition qui prend la forme de potentiels synaptiques inhibiteurs rapides et globaux au neurone, se présente sous deux aspects : elle est tout d'abord divisive lors de l'ouverture des canaux ioniques, c'est-à-dire pendant quelques millisecondes, puis soustractive comme la conductance de la membrane augmente.

*où $e = \exp(1)$ évalue $\sum_{1 \leq k \leq N} \frac{1}{k!}$

Ici, nous nous intéresserons particulièrement à ces premières millisecondes où la conductance membranaire peut alors jusqu'à tripler [BorgGraham, 1999] divisant alors d'autant l'efficacité des PPS excitateurs. À l'instar de Delorme et Thorpe [2001a] qui ont utilisé le logiciel de simulation NEURON, nous allons implanter un modèle de neurone décodant un signal asynchrone grâce à l'inhibition divisive, mais en utilisant un formalisme explicite simplifié.

Modèle de neurone sensible à l'asynchronie

Plus précisément, imaginons alors un neurone i avec N synapses j sur lesquelles il reçoit des impulsions qui modifient alors son potentiel, modélisé par les équations LIF-CS (voir Sec. 3.1.1 et Éq. 3.1.9) et que l'on peut donc paramétrer par un vecteur de poids et le seuil. On définit alors l'inhibition divisive comme un processus qui va intégrer avec une même efficacité synaptique les impulsions, donc les compter dynamiquement, pour générer un courant ionique (dont le potentiel de repos est proche du potentiel excitateur du neurone) et dont l'effet sera diviseur sur l'efficacité du courant excitateur. L'efficacité de la synapse j sera donc transformée en

$$g^*_{ij} = g_{ij}/g_{\mathtt{shunt}} \tag{3.3.1}$$

avec

$$\frac{d}{dt}g_{\mathtt{shunt}}(t) = \frac{1}{\tau_{\mathtt{syn}}}.(1 - g_{\mathtt{shunt}}(t)) + \alpha.g_{\mathtt{shunt}}(t).\sum_j S_j \tag{3.3.2}$$

où α est le taux d'inhibition par impulsion entrante et qu'on réglera suivant les données neurophysiologiques.
On peut approcher qualitativement la réponse d'un tel modèle si l'on considère une vague d'impulsions où chaque synapse reçoit une et une seule impulsion au temps t_j. En effet, l'inhibition divisive grandira en suivant une exponentielle décroissante du nombre d'impulsions arrivées sur le neurone et on pourra écrire comme dans Sec. 3.1.3 que l'activité excitatrice du neurone sera

proportionnelle à :

$$g_i^* = \sum_j g_{ij}^* \sim \sum_j w_{ij}.[\Delta(t - t_{ij}).\exp(-\sum_j \Delta(t - t_{ij}))] \tag{3.3.3}$$

Il ressort de cette expression que comme $\sum_j \Delta(t - t_{ij})$ est une mesure explicite du rang au temps t, le courant intègre des valeurs proportionnelles à l'efficacité de la synapse mais modulées par le rang. On en déduit donc que qualitativement l'activité sera d'autant plus forte que la corrélation entre les poids et le vecteur décroissant avec le rang sera fort, c'est-à-dire que les poids seront dans le même ordre que l'ordre temporel des impulsions.
On a simulé l'évolution de ce modèle suivant les mêmes méthodes que précédemment (voir Sec. 3.1.1) mais avec maintenant un front d'impulsions codé de façon asynchrone, généré rythmiquement avec des délais modulées dans le temps suivant une sinusoïde (ce front simule par exemple la réponse d'une rétine monodimensionnelle à une "mire" sur laquelle le contraste varie périodiquement). Ce même signal est permuté à travers les fibres et génère plusieurs fronts différents d'impulsions suivant la même modulation sinusoïdale. On observe alors (voir Fig. 3.3.1) que le neurone adapté au décours temporel des impulsions, *i.e.* dont les poids sont en corrélation avec un ordre particulier d'arrivée des impulsions, génère une impulsion à la présentation du signal, alors que ce même signal permuté aléatoirement ne génère pas de sortie. On a donc construit grâce à la *shunting inhibition* un modèle biologique de décodage de l'asynchronie par un codage temporel par l'ordre d'activation. De plus, on peut adapter la règle d'apprentissage précédente (voir Sec. 3.1.2), en modifiant la fenêtre d'apprentissage en fonction du nouveau mécanisme d'inhibition divisive que nous avons introduit. Ceci modifie particulièrement la forme de cette fenêtre dans le sens où dans le cas d'une relation causale, l'augmentation de poids varie comme un PPS. Les résultats montrent que la distribution de poids converge pour un signal asynchrone vers les valeurs d'inhibition respectives.

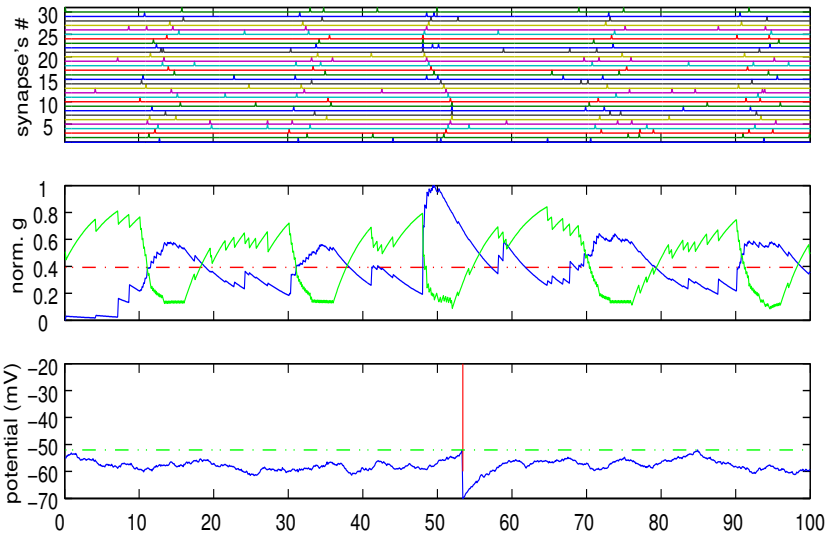

FIG. 3.3.1 – *Modèle de neurone détecteur d'asynchronie.* **(Haut)** On applique sur 30 fibres pendant 100 millisecondes le même signal impulsionnel modulé en latence par une sinusoïde mais permuté aléatoirement entre les fibres. Ce motif est présenté régulièrement à des rythmes proche du rythme γ (50 Hz). Le motif est ordonné pour la figure au temps 50 ms. **(Milieu)** Ces vagues d'activation bruitées génèrent au soma du modèle de neurone un courant excitateur (trait bleu-sombre) modulé par la résistance d'un courant diviseur de *shunting inhibition* (trait vert-clair) et qui sert de modulation en fonction du rang. Chaque vague produit une inhibition équivalente car les signaux sont équivalents à une permutation près. **(Bas)** Comme les poids sont adaptés à un tir dans l'ordre correspondant au motif central à $t = 50\,\mathrm{ms}$, le neurone est plus excité et produit éventuellement une impulsion : on a alors décodé le signal asynchrone.

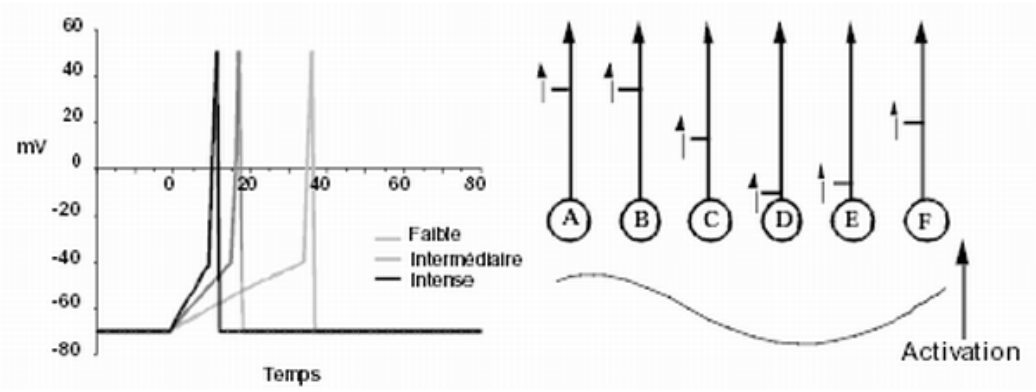

FIG. 3.3.2 – *Principe du codage d'une valeur par son rang.* **(A)** Si l'on considère un neurone intégrateur excité par des valeurs analogiques constantes, il tirera d'autant plus vite que cette valeur est forte. **(B)** Un vecteur de valeurs analogiques peut être codé en émettant de façon parallèle des impulsions uniques d'autant plus rapidement que cette valeur est forte. On pourra ainsi extraire une information sur ce vecteur en observant le vecteur des rangs des impulsions soit simplement dans ce cas $B > A > F > C > E > D$. (source [Van Rullen, 2001])

Propriétés et principes du codage par rang

Cet exemple de mécanisme neuronal présente donc une possibilité offerte pour décoder en plus de la synchronie l'information temporelle asynchrone codée par l'ordre d'activation des PAs et qui pourrait implanter les mécanismes du modèle de rétine (voir Sec. 3.2). Comme il a été proposé par Thorpe et Imbert [1989], ce code peut alors correspondre à un type de processus utilisé pour le traitement ultra-rapide de l'information visuelle. On peut alors dégager un principe fondamental du codage par rang : l'information est transmise en avant, de façon asynchrone et entre deux couches neurales, l'information analogique est codée temporellement puis décodée par l'ordre d'activation. Par conséquent, on pourra implanter un système plus simple en écrivant explicitement la modulation suivant le rang et en construisant un modèle de type Intègre-et-Tire utilisant le rang en entrée (voir Fig. 3.3.2).

De nombreuses études menées dans le laboratoire de Simon Thorpe ont montré les performances d'un tel codage par le rang [Thorpe et Gautrais, 1998] par rapport au codage fréquentiel [Gautrais, 1997; Gautrais et Thorpe,

1997], au codage par synchronie [Delorme, 2001, p.162] ou encore à d'autres codages temporels comme l'intervalle inter-impulsionnel [Van Rullen et Thorpe, 2001]. Ces études ont été complétées en collaboration avec M. Samuelides par des études mathématiques de ce code. On a notamment étudié son rapport avec les statistiques non-paramétriques [Barbe et Hallin, 1995], comme le test de Wilcoxon mais aussi analysé ses performances pratiques grâce aux travaux de Mélis [1995] mais aussi sur l'apprentissage d'orientations d'E. [Veneau, 1996] et enfin l'apprentissage supervisé de motifs [Samuelides et al., 1997] et de textures de J. Vaccaro et D. Gourion [Vaccaro et al., 1998].
Aussi, il est important d'évoquer le caractère général de ce codage qui n'apparaît pas dans sa formalisation. En effet, si pour se démarquer des nombreux codes utilisant l'information de population ou de fréquence, le codage par rang se concentre sur une propagation en avant avec une impulsion par neurone, le quanta de l'information neurale, ici l'impulsion, peut dans le cadre du modèle être étendu à des signaux neuronaux différents. Par exemple, des bouffées d'impulsions (ou *bursts*) sont caractéristiques de certains neurones ; mais on peut aussi imaginer que des "petites" assemblées de neurones qui réagissent en synchronie puissent ainsi exhiber des réponses plus robustes.

3.3.2 Étude théorique du codage par rang

Nous allons maintenant expliciter et étudier ce modèle pour montrer la richesse d'un calcul basé sur les rangs d'activation.

Formalisation mathématique du codage par rang

On se placera dans cette section dans le cas où l'information est simplement codée par une vague d'impulsions, avec au plus une impulsion par fibre, et qu'on l'on peut donc déterminer cette vague par la liste ordonnée $o(r)$ des impulsions, où r est le rang de l'impulsion et $o(r)$ "l'adresse" de la fibre correspondante. À l'inverse, on peut écrire le rang d'une fibre o par $r(o)$. On

écrira alors le potentiel du neurone dans notre modèle pour $r \geq 1$ comme :

$$\begin{cases} V(r) = V(r-1) + w_{o(r)}.Mod(r) \\ \text{Impulsion et arrêt pour } V(r) > V_{\texttt{seuil}} \end{cases} \tag{3.3.4}$$

où, à l'image de l'inhibition divisive dans la section précédente, on définit une fonction de modulation dépendant du rang par une fonction décroissante. Par simplicité, le neurone est initialisé à $V(0) = 0$ et $Mod(0) = 1$ et comme dans les modèles SRM et LIF, le neurone génère une impulsion lorsqu'il atteint un seuil $V_{\texttt{seuil}}$, et est alors inhibé totalement (il ne peut donc tirer qu'une seul impulsion). On en déduit immédiatement que le potentiel au rang t est donné par

$$V(t) = \sum\nolimits_{1 \leq r \leq t} w_{o(r)}.Mod(r) \tag{3.3.5}$$

On en déduit alors que comme on contrôle la somme des poids, on peut fixer de façon arbitraire $V_{\texttt{seuil}} = 1$. De plus, si on note o_w le vecteur d'ordre des poids tel que le vecteur des poids $w_{o_w(.)}$ est maintenant rangé par ordre décroissant,

$$S_{\texttt{min}}(t) \leq V(t) \leq S_{\texttt{max}}(t) \tag{3.3.6}$$

avec si on note $r_{\texttt{max}}$ le rang maximal (*i.e.* le nombre total de fibres),

$$\begin{cases} S_{\texttt{min}}(t) = \sum_{1 \leq r \leq t} w_{o_w(r_{\texttt{max}}-r+1)}.Mod(r) \\ S_{\texttt{max}}(t) = \sum_{1 \leq r \leq t} w_{o_w(r)}.Mod(r) \end{cases} \tag{3.3.7}$$

C'est à dire qu'à chaque rang le maximum (resp. le minimum) est réalisé pour une vague d'impulsions arrivant dans l'ordre (resp. l'ordre inverse) des poids (voir Fig. 3.3.3).
Cette mesure étend donc les résultats des statistiques non-paramétriques, c'est-à-dire des tests statistiques dont le résultat ne se base pas sur une évaluation de la densité de probabilité des variables. On comprend que ces méthodes sont particulièrement adaptées au signal neuronal puisque la densité de probabilité d'arrivée des impulsions dépend de chaque stimulus et de

son traitement à des niveaux antérieurs et le résultat y sera ici indépendant. L'adéquation de ces tests à la tache sera alors contrôlée par les fonctions de poids et de modulation.
En effet, la fonction de modulation va jouer un rôle majeur car elle va déterminer l'importance relative des ordres d'arrivée, une modulation décroissant très rapidement privilégiant alors une information concentrée sur un nombre restreint d'impulsions, le code sera alors dit *épars* (en anglais *sparse*). Aussi comme il a été évoqué dans [Perrinet, 1999] et appliqué dans [Van Rullen et Thorpe, 2001], on peut choisir la modulation de telle sorte qu'elle corresponde en moyenne à la valeur analogique codée par le rang : c'est la méthode utilisée à la Sec. 3.2.2 qui nous a permis de coder par leur rang les valeurs absolues des coefficients des ondelettes. Enfin, le décodage de cette mesure se fera alors par la génération d'un impulsion lorsque le potentiel dépassera un certain seuil. La réponse est donc dynamique puisqu'elle code de façon temporelle la corrélation mais aussi car cette information est ensuite transmise de façon asynchrone au reste du réseau.

Applications du théorème central limite permutationnel

Étudions dans un premier temps la statistique du potentiel si l'on abstrait le processus de génération de l'impulsion. Le *théorème central limite permutationnel* [Barbe et Hallin, 1995] montre* alors que la loi de probabilité du potentiel pour des vecteurs de rang aléatoires est approximaitvement gaussienne. Or on montre d'autre part que les moments statistiques du potentiel peuvent être facilement déduits de ceux du poids et de la modulation. Pour la moyenne et la variance au rang t, si l'on note Mod_t la fonction de modulation tronquée aux t premières valeurs et valant 0 ensuite,

$$\begin{cases} \langle V(t) \rangle = r_{\texttt{max}} . \langle w \rangle \langle Mod_t \rangle \\ \sigma^2_{V(t)} = (r_{\texttt{max}} - 1) \sigma^2_w \sigma^2_{Mod_t} \end{cases} \tag{3.3.8}$$

*sous la condition de Noether sur les poids que l'on supposera vérifiée [Gautrais, 1997], ce qui est vrai en pratique pour des modulations décroissant rapidement.

et ces calculs s'étendent aussi aux calculs des moments d'ordre arbitraire[†]. On peut donc estimer pour tout neurone —étant donnés ses poids et sa fonction de modulation— la densité de probabilité du potentiel à chaque impulsion, c'est-à-dire à chaque rang t (voir Fig. 3.3.3). Ce résultat permet alors d'estimer pour une sensibilité désirée du neurone le seuil en fonction des poids et de la fonction de modulation en utilisant les quantiles de la loi de répartition normale [Perrinet, 1999].

Si l'on tient maintenant compte du mécanisme impulsionnel, comme on sait estimer à chaque pas de temps t (*i.e.* à chaque rang) la loi de probabilité du potentiel linéaire par une gaussienne, on peut estimer la loi de probabilité $p(t)$ du rang de tir du neurone, c'est-à-dire prévoir le comportement impulsionnel du neurone à codage par rang. On va pour cela assimiler ce problème à l'évaluation du temps d'arrêt du processus aléatoire (voir [Burkitt et Clark, 1999; Horn et Levanda, 1999]) décrit par l'Éq. 3.3.4. En effet, soit $f(t)$ la probabilité que le potentiel non arrêté $V(t)$ soit supérieur au seuil $V_{\texttt{seuil}} = 1$ au rang t.

$$f(t) = P(V(t) > 1) \tag{3.3.9}$$

et que l'on sait évaluer grâce au théorème précédent et aux moments donnés par Éq. 3.3.8 de la moyenne et de la variance en fonction du rang. Au rang t, la probabilité $p(t)$ que le neurone tire à t, est la probabilité qu'il n'a pas déjà tiré et qu'il atteint le seuil au temps t. Comme $V(t)$ est croissant, on a simplement

$$\left\{ \begin{array}{rcl} p(t) & = & P(V(t) > 1) - P(V(t-1) > 1) \\ & = & f(t) - f(t-1) \end{array} \right. \tag{3.3.10}$$

On déduit finalement des moments des poids et de la modulation la densité de probabilité de tir d'un neurone pour des entrées aléatoires. Ce calcul permet rétroactivement de prévoir ces paramètres pour obtenir une réponse

[†]Le principe du calcul est d'utiliser les symétries dans l'ensemble Π des permutations. Ainsi, pour la moyenne $\langle V(t) \rangle = \frac{1}{n!} \sum_{p \in \Pi} \sum_{1 \leq i \leq t} (w_{p(i)} Mod(i)) = \sum_{1 \leq i \leq t} Mod(i) \frac{1}{n!} \sum_{p \in \Pi} w_{p(i)} = [\sum_{1 \leq i \leq t} Mod(i)]. \langle w \rangle$. Le calcul est plus long pour la variance, mais il utilise le même principe.

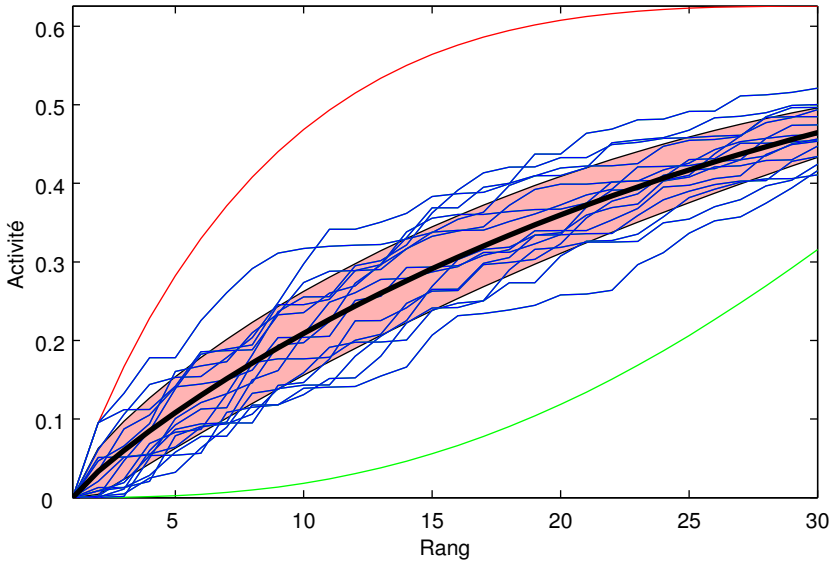

FIG. 3.3.3 – *Etude statistique de l'activité d'un neurone à codage par rang.* On a simulé des intégrations neuronales linéaires sur 30 synapses par un codage par rang pour des entrées aléatoires (et donc des permutations aléatoires de l'ordre d'activation des synapses). Les courbes extrêmes correspondent aux valeurs théoriques maximales et minimales du potentiel, alors que le trait gras correspond à sa moyenne et la zone pleine à un écart d'un écart type de la moyenne. Les autres traits correspondent à des réalisations du potentiel pour des vecteurs aléatoires. On observe en pratique que la densité de probabilité du potentiel à chaque rang est donné avec une bonne approximation par le théorème central limite permutationnel.

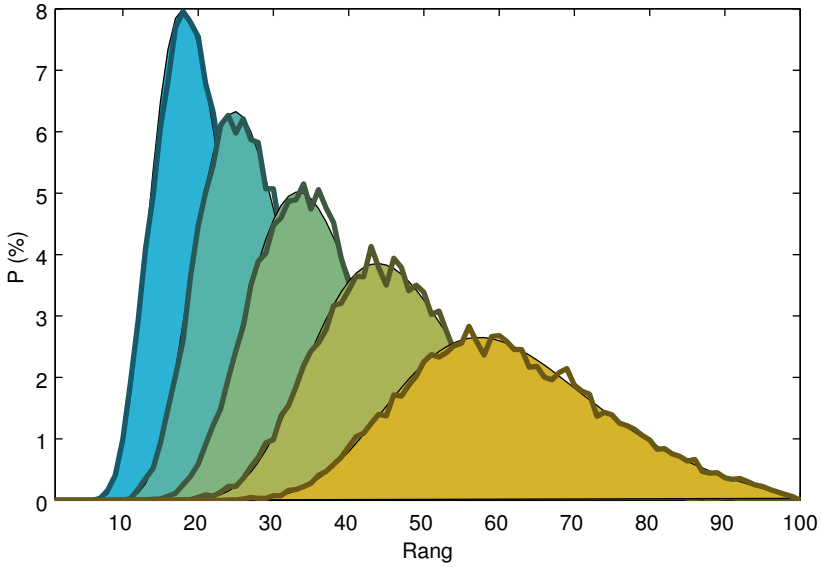

FIG. 3.3.4 – *Estimation de la densité de probabilité du temps de tir d'un neurone à codage par rang.* Grâce à l'estimation du potentiel d'un neurone linéaire (voir Fig. 3.3.3), on peut estimer la densité de probabilité de tir du neurone impulsionnel. On présente, pour un neurone avec 100 synapses, l'estimation théorique de la densité de probabilité (en zone pleine) avec (en traits gras) la densité mesurée pour des vecteurs de rang aléatoires. On a fait varier le seuil (*i.e.* la somme des poids) pour montrer que cette méthode permet de régler explicitement plusieurs stratégies de codage : rapide avec un seuil bas, ou plus lent et étalé avec un seuil haut.

---

désirée. Par exemple, si l'on veut que le neurone tire en moyenne après un certain rang, on pourra en déduire le seuil adéquat en réglant la somme des poids. Ce résultat a été simulé (voir Fig. 3.3.4) et montre que notre hypothèse est suffisante pour prévoir la densité de tir d'un neurone avec précision. On pourra donc aussi calculer les paramètres pour optimiser la réponse du neurone comme l'entropie de la densité de probabilité de tir pour promouvoir lors d'une propagation sur plusieurs couches neurales la meilleure transmission de l'information. En particulier, on évite ainsi au maximum les impulsions de rangs égaux dont le traitement peut poser des problèmes et qui sont très sensibles au bruit.

Apprentissage

Les principes d'apprentissage par plasticité à modulation temporelle relative (voir Sec. 3.1.2) peuvent être de même prolongés spécifiquement au codage par rang. Comme évoqué, la fenêtre temporelle doit être modifiée pour que l'adaptation des poids soit fonctionnelle. Par analogie, la règle que nous utilisons va donc utiliser non la différence de temps entre les impulsions pré- et post-synaptiques, mais directement leur rang. De manière qualitative, on va donc renforcer positivement la valeur des poids des synapses en fonction de leur rang vers la valeur correspondante de modulation jusqu'au tir du neurone puis inhiber les autres synapses, ce qui correspond expérimentalement aux observations biologiques [Perrinet et al., 2001]. Cette formalisation permet alors d'établir une règle simple d'apprentissage de type hebbien qui présente par sa structure des règles homéostatiques explicites, comme la conservation de la somme des poids.

Dans un article compagnon [Delorme et al., 2001b], cette règle a été appliquée dans un modèle de l'aire visuelle V1 sur laquelle on fait converger l'information rétinienne des contrastes. Les stimuli étaient constitués de 2790 images de taille 364×244 en niveaux de gris (sur 8bit) dans lesquellles on extrait des sous-images de 50×50. Celles-ci étaient codées sur une échelle par une rétine constituée de cellules ON et OFF puis décodées grâce à un codage par rang sur la deuxième couche. Les neurones de la deuxième couche intègrent alors l'information suivant le modèle que nous avons décrit et avec une fonction de modulation fixée. Les seuils des neurones (soit les sommes des poids) varient au cours de l'apprentissage pour adapter la sensibilité des neurones à leur convergence grâce à des techniques statistiques.

Des simulations de ce modèle ont alors montré que les poids des vecteurs convergeaient vers des filtres localement sensibles à la direction, comme cela s'observe dans les colonnes corticales de V1 [Hubel et Wiesel, 1959]. En fait, ce résultat est semblable à d'autres travaux conduisant à l'émergence de filtres similaires [Olshausen et Field, 1996] et montre que le codage par rang peut de la même manière prédire la formation de structures émergentes. On peut remarquer que la loi d'apprentissage que nous appliquons va faire émerger

des filtres tels que leur distribution de sortie soit sur-gausienne, c'est-à-dire que pour la même variance, plus de sorties sont proches de la moyenne.

3.3.3 Applications

Sélectivité ultra-rapide à la direction

L'apprentissage évoqué dans la section précédente suggère d'étendre le modèle de la rétine de Van Rullen et Thorpe [2001] à l'aire V1 et d'étudier la possibilité d'un codage par rang de l'image dans cette aire. En effet, des mesures neurophysiologiques suggèrent que les neurones de V1 sont sensibles à travers leur fréquence de décharge à des orientations précises mais aussi que de la même manière, ils tirent d'autant plus rapidement que l'orientation présentée est proche de l'orientation préférée. La sélectivité à l'orientation est donc aussi codée dans la latence de tir des neurones [Celebrini et al., 1993] et pour étudier la possibilité de détecter des orientations dans une image, nous allons construire un modèle simple de ce système.
Ce modèle va reprendre le modèle de rétine comme précédemment (mais dans un but de simplification avec une seule échelle) et nous allons simuler de la même manière un neurone de l'aire V1 caractérisé par son champ récepteur, sa taille et son vecteur de poids. Dans un premier temps, nous définirons l'entrée comme des images de contours à différentes orientations centrés sur le champ récepteur du neurone. En définissant un modèle de contour, on peut alors calculer le vecteur de poids comme la réponse modulée d'une orientation (que nous appellerons par référence 0°). L'activité du neurone de V1 sera alors mise à jour à chaque impulsion arrivant sur la couche suivant une fonction de modulation à décroissance géométrique que nous changerons pour en étudier l'influence. Nous noterons alors si le neurone dépasse un seuil, auquel cas il émettra une impulsion.
Pour intégrer la plausibilité biologique avec l'efficacité du codage par rang, nous avons représenté les impulsions suivant leurs latences pour plusieurs valeurs de contraste alors qu'elles sont décodés suivant un codage par rang. Alors, si l'on choisit des paramètres vraisemblables pour le délai de traitement et de délai jusqu'à V1, la réponse des neurones de cette simulation correspond

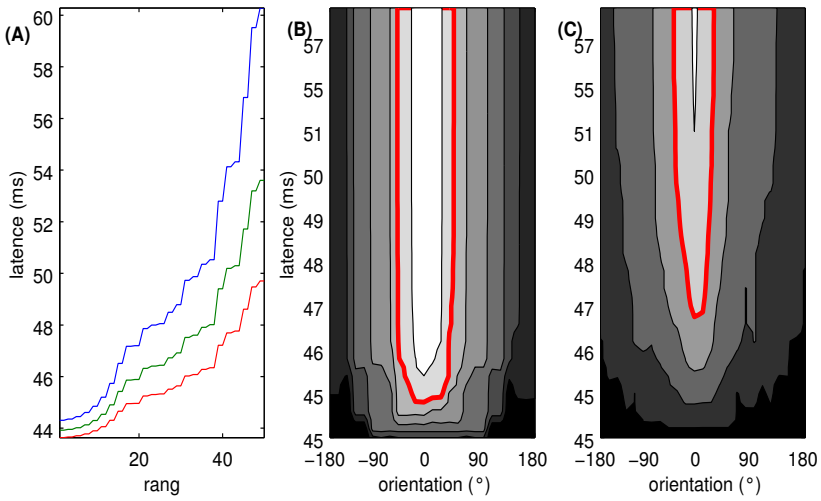

FIG. 3.3.5 – *Sélectivité Ultra-rapide à l'orientation dans V1.* On a simulé un neurone auquel on présente l'image de barres orientées. **(A)** Latence équivalente de décharge en fonction du rang pour différents contrastes de l'entrée. Le neurone de sortie est sélectif à une orientation préférée (0°). **(B et C)** Influence du seuil sur la largeur de la courbe de sensibilité : les niveaux de gris indiquent la réponse du neurone pour le choix de différents seuils, la forme de la courbe de sensibilité est sensiblement similaire mais plus rapide et large pour des seuils bas. **(B vs. C)** Influence de la fonction de modulation. On a choisi en (C) une fonction décroissant plus lentement (fonction géométrique de raison 0.9 au lieu de 0.95 pour (B)) : la réponse est plus lente, mais la courbe de sensibilité est plus resserrée et précise (traits gras).

aux observations de Celebrini et al. [1993] et montre aussi que la largeur de la courbe de sensibilité est bien indépendante du contraste, ce qui est une conséquence directe du choix du codage. On remarque aussi que cette forme de la courbe de sensibilité en fonction de l'orientation peut être modifiée par plusieurs facteurs. Un seuil plus bas correspond à une latence plus rapide mais une réponse moins précise et robuste. Complémentairement, si l'on "ralentit" la décroissance de la modulation, on observe alors qu'on peut rétrécir la largeur de la courbe de sensibilité.

Détection de la meilleure concordance dans une image

Dans le cadre du traitement visuel, on voit que le codage par rang par la richesse des diverses représentations impulsionnelles et la flexibilité du décodage est particulièrement adapté à la détection de concordance avec des dictionnaires de motifs dans l'image. En effet, si l'on peut comprendre le traitement rétinien et dans V1 en termes d'analyse fréquentielle en sous-bandes d'échelle et d'orientations, on peut aussi caractériser ce traitement comme la recherche de motifs appris qui vont hiérarchiquement correspondre aux points, orientations, lignes, courbures. On peut alors définir plus exactement l'algorithme de codage par rang pour exploiter au mieux la caractère progressif de cette recherche.
En effet, nous avons défini jusqu'à maintenant le rang de façon globale à un ensemble de neurones, mais on comprend que dans une image par exemple, si le contraste change de façon continue, par la position de l'éclairage par exemple, il est désirable que la réponse des neurones reste la même et donc qu'elle soit qualitative au lieu d'être quantitative. Il est alors nécessaire d'utiliser le *rang local*, c'est-à-dire le rang calculé pour un neurone, comme il a été suggéré dans le modèle biologique.
Nous avons simulé une telle application pour une recherche de motifs dans une image dont le contraste transversal varie suite à une erreur de numérisation suivant une rampe. Les résultats montrent que si l'on utilise le rang global, des motifs incorrects peuvent être propagés avant des motifs corrects mais peu contrastés (voir Fig. 3.3.6-Milieu). Suivant l'application, on pourra plutôt choisir le *rang local* : les motifs correspondent à des réponses similaires indépendamment du contraste local (voir Fig. 3.3.6-Bas). Cette dernière propriété est commune dans le système visuel et permet d'étendre la gamme dynamique de notre perception. Toutefois, si ce processus est évident à implanter dans un système de calcul parallèle, il est prohibitif dans un calculateur de type von Neumann.

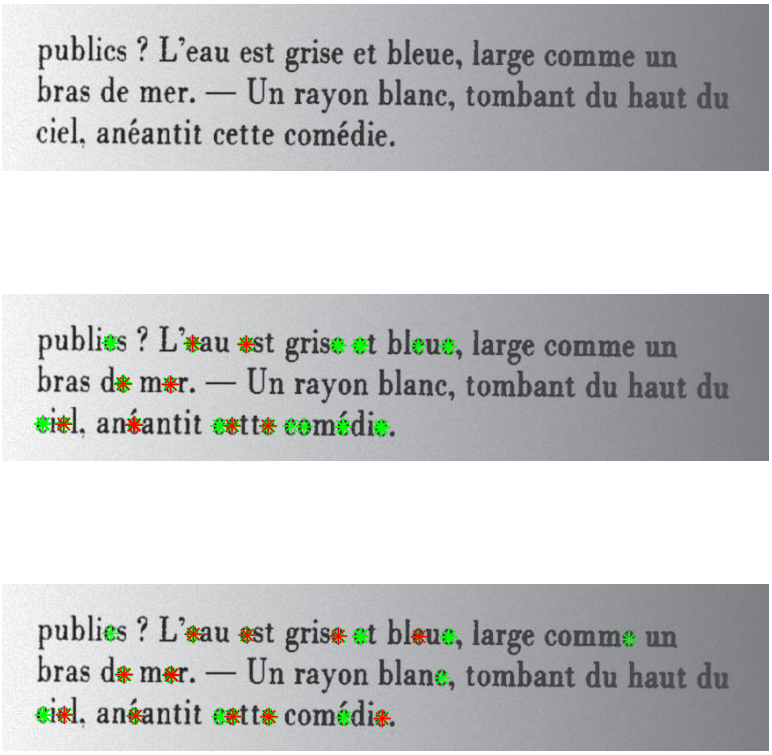
publics ? L'eau est grise et bleue, large comme un
bras de mer. — Un rayon blanc, tombant du haut du
ciel, anéantit cette comédie.
publics ? L'eau est grise et bleue, large comme un
bras de mer. — Un rayon blanc, tombant du haut du
ciel, anéantit cette comédie.
publics ? L'eau est grise et bleue, large comme un
bras de mer. — Un rayon blanc, tombant du haut du
ciel, anéantit cette comédie.

FIG. 3.3.6 – *Recherche de concordance.* Le codage par rang est particulièrement adapté à la recherche de motifs appris dans l'image. (**Haut**) On a utilisé le rang global pour calculer le potentiel linéaire d'un neurone sélectif à un motif, ici la lettre "e" en chaque point de l'image (**Milieu**) en émettant progressivement des impulsions (étoiles rouge puis vertes). Toutefois, il existe une variation de contraste suivant une rampe transversale et des neurones répondent à des motifs incorrects mais plus contrastés. (**Bas**) On a alors utilisé le rang local pour déterminer le potentiel des neurones : la réponse est alors indépendante du contraste mais varie seulement avec le rapport signal sur bruit. La réponse impulsionnelle est meilleure, et ily a moins de fausses détections (source du texte : A. Rimbaud, *Révélations*, BNF.).

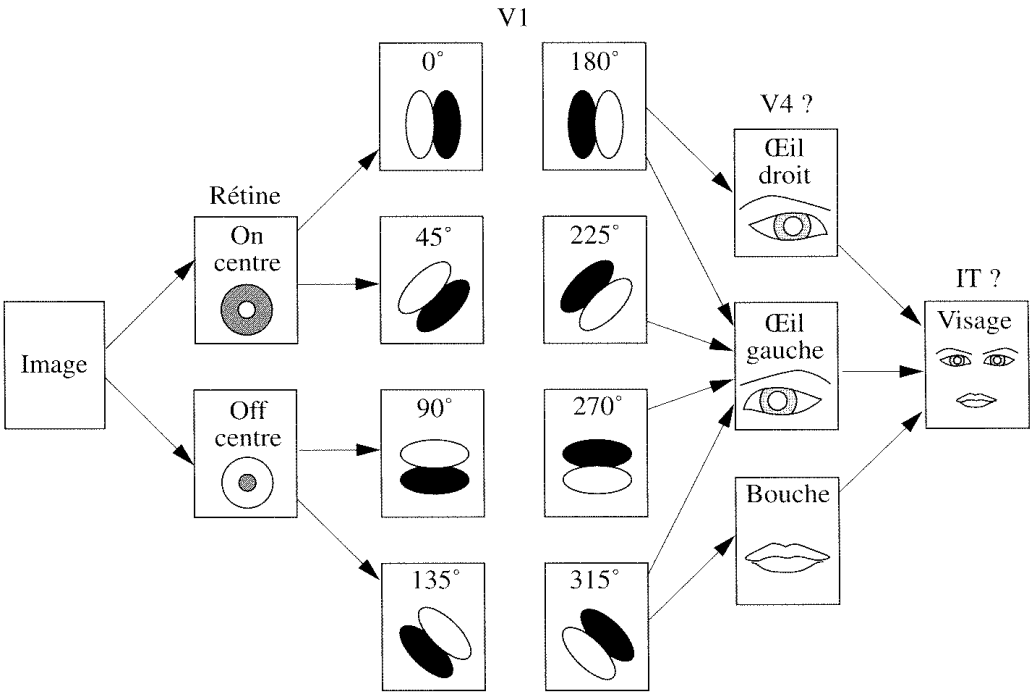

FIG. 3.3.7 – *Architecture de* SPIKENET. L'image est codée de façon asynchrone suivant des chemins divergents de la rétine sur l'aire V1 puis converge vers une carte similaire à V4. Cette carte détecte des parties de l'image (yeux, bouche) puis enfin l'activité dans la dernière carte correspond à la localisation d'un visage (source [Van Rullen et al., 1998]).

Spikenet

Finalement, on peut imaginer un système entier basé sur ces principes pour implanter un processus visuel complet de catégorisation ultra-rapide. C'est ce qui a été réalisé par l'équipe du Cerco [Van Rullen et al., 1998] pour la détection de visages. Dans ce réseau de neurones, l'architecture est composée de cartes neuronales successives correspondant à : la rétine avec des filtres ON et OFF détectant le contraste, l'aire V1 détectant les orientations, l'aire V4 détectant des formes simples comme les parties du visage (œil, bouche) et l'aire PIT comportant des neurones sensibles à des visages (voir Fig. 3.3.7).

Les calculs seront donc totalement asynchrones, et se propageront depuis la

rétine jusqu'à la dernière carte en mémorisant pour chaque carte les activités des neurones et en propageant les PAs émis depuis une carte intermédiaire ou depuis la rétine. En pratique, le nombre prohibitif de neurones et de synapses dans cette simulation par rapport à la taille actuelle des mémoires vives des calculateurs oblige à avoir recours à des implantations virtuelles, c'est-à-dire que les paramètres d'un neurone vont être donnés par des méta-paramètres comme par exemple pour les neurones de V1 leur localisation, taille et orientation. Les neurones de V4 et IT sont appris de manière supervisée en découplant les différentes parties de l(image que l'on désire apprendre "à la main".

Cette architecture illustre parfaitement la performance du codage par rang dans le cadre de la catégorisation ultra-rapide, autant par l'architecture parallèle et asynchrone du calcul que par sa robustesse, par exemple aux changements de contrastes et de façon équivalente au bruit [Delorme et Thorpe, 2001c]. Ainsi, dans le cadre de la start-up SPIKENET TECHNOLOGIES, ces algorithmes sont utilisés pour des applications industrielles. Ces technologies peuvent être, grâce à leur structure, implantés sur des systèmes distribués comme des *clusters* (c'est-à-dire une réseau de calculateurs dédié au calcul), des cartes multiprocesseurs. Cette représentation se rapproche de la Représentation par l'Adresse des Événements (Address-Event Response, ou *AER*, [Deiss et al., 1999]) et peut donc être de façon simple transposée à des cartes de type VLSI. Ces fondations ouvrent donc la voie à l'implantation de systèmes de vision complets dont la rapidité rivalise avec les systèmes biologiques.

Chapitre 4

Vers un code dynamique épars

> Il faudrait faire en sorte que tout soit aussi simple que possible mais pas plus simple
>
> Albert Einstein,
> cité par Marvin Minsky, in *Society of Mind*

4.1 Pour un modèle d'interactions latérales

Après avoir présenté un modèle complet de codage rétinien (Sec. 3.2) et un code impulsionnel basé sur les rangs d'arrivée des impulsions (Sec. 3.3), nous allons voir maintenant pourquoi il est nécessaire de proposer un modèle d'interactions latérales pour implanter un code visuel efficace.

4.1.1 Densité du codage et représentation sur-complète

Quelle représentation visuelle par impulsions ?

Tout d'abord, le codage par une décomposition en ondelettes se heurte à des problèmes de vraisemblance biologique. Comme le remarquait Van Rullen et Thorpe [2001], le codage sur une grille dyadique n'imite pas la répartition des tailles des champs récepteurs des cellules ganglionnaires dans la rétine et

de plus les filtres correspondants ne sont pas en général indépendants [Meister et Hosoya, 2001]. Il est encore plus important de noter que les champs récepteurs biologiques ne vérifient pas en général la condition de la formule d'inversion qui permet la reconstruction (voir Éq. 3.2.8) et plus généralement que la transformée rétinienne que nous avons étudié ne s'appliquera à la reconstruction (on dit que la base des filtres rétiniens ne forme pas un cadre serré ou *tight frame* [Salinas et Sejnowski, 2001]). Aussi, ces difficultés se traduisent par les contraintes qu'imposent la condition de reconstruction pour définir une base adaptée de filtres dans l'architecture du modèle.

D'autre part, le résultat du codage rétinien que nous avons étudié est instable par translation. Ainsi, alors que les neurones du système visuel, et en particulier de la rétine, semblent avoir des réponses continues pour de petites transformations usuelles de l'image (translation, agrandissement, rotation), les neurones d'une transformée en ondelettes sur une grille dyadique peuvent avoir des comportements différents (voir Fig. 4.1.1). Une conséquence supplémentaire est que les processus de reconnaissance et d'apprentissages basés sur ce codage seront d'autant plus complexes à implanter s'ils utilisent une telle représentation.

Aussi, si, pour résoudre ce problème, on augmente le nombre de filtres en insérant des tailles intermédiaires entre les cartes dyadiques, alors le code résultant devient redondant. En effet, cette stratégie correspondra simplement à une discrétisation plus fine de la transformée en ondelettes continue, et en augmentant le nombre de filtres, on ne change pas la décroissance des coefficients de la transformée continue. De cette sorte, les coefficients de la transformation discrète (et indirectement la convergence de la reconstruction) sera d'autant plus redondante et lente, *i.e.* il faudra alors un nombre proportionnellement aussi grand de PAs pour obtenir une reconstruction équivalente, ce qui ne semble pas être ni une stratégie satisfaisante ni le comportement observée dans les aires visuelles primaires [Vinje et Gallant, 2000].

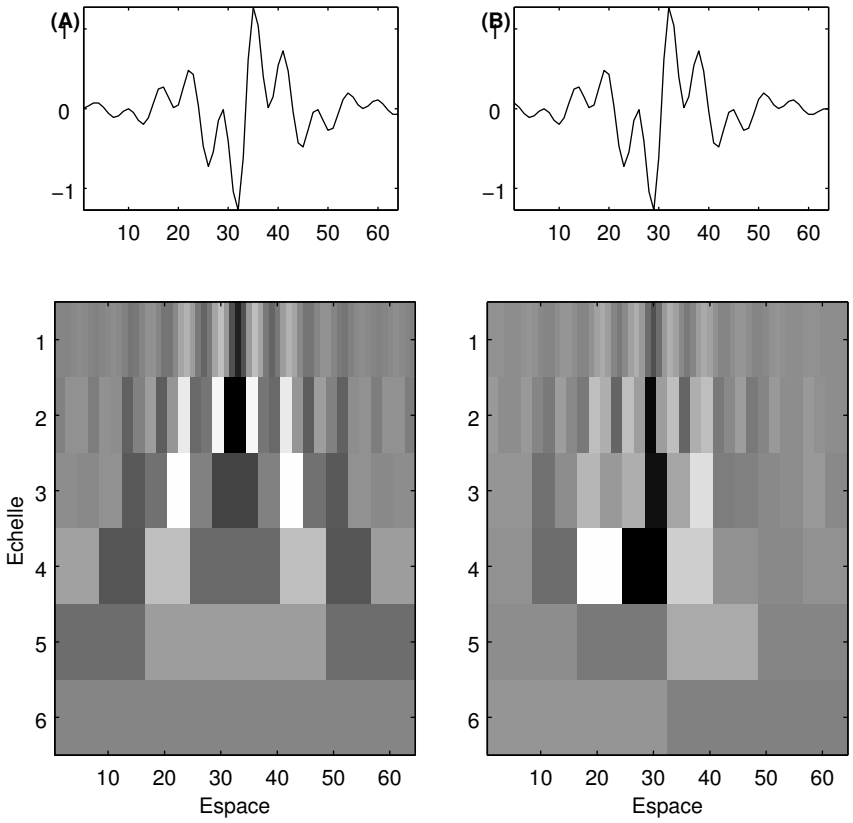

FIG. 4.1.1 – *Instabilité de la représentation par ondelette.* On a représenté (**Bas**) la transformée avec des ondelettes de Haar mono-dimensionnelle sur une grille dyadique d'un signal (**A**) sur 64 points et (**A**) de sa translation de 3 points (**Haut**). Les "signatures" correspondant à la transformation sont instables à une petite variation en translation et sont inadaptées à la représentation d'un signal de type visuel (les valeurs grises sont proches de 0, les activités extrémales sont blanches et noires).

Vers un codage impulsionnel épars

Une alternative est alors de trouver à l'image du système visuel une *représentation éparse** de l'image [Olshausen et Field, 1998], comme ce qui est observé dans le système visuel [Baddeley, 1997]. Par exemple, nous avons vu que les neurones de IT peuvent répondre à des signaux très spécifiques (jusqu'au neurone "Bill Clinton") mais qu'ils pouvaient répondre de façon continue à des transformations courantes pour l'information visuelle (par exemple liées aux déplacements du sujet). Cette représentation à mi-chemin entre une représentation dense distribuée (comme ci-dessus) et une représentation locale (le neurone "grand-mère") va se caractériser dans une assemblée par un faible taux d'activité à travers les neurones. Typiquement, la représentation dans le temps par un code impulsionnel épars sera alors constituée à un moment donné d'un faible nombre relatif d'unités activées de façon cohérente.
Dans le cadre du codage de l'information visuelle, ce code doit pouvoir donner une représentation stable pour les transformations usuelles, c'est-à-dire que la représentation de la transformée d'une image peut se rapprocher d'une transformée de la représentation de l'image (à l'opposé donc de ce que l'on a vu a la Fig. 4.1.1). Ainsi, par exemple, on peut souhaiter que la représentation d'une image grossie puisse se déduire d'une transformation simple de la représentation originale. Un avantage computationnel immédiat est la généralisation de cette représentation (d'un objet par exemple) depuis des images prototypes à l'ensemble d'images générées par les compositions de ces transformations et par exemple à l'apprentissage d'un tel objet depuis un nombre limité de présentations. Un autre avantage est que le codage épars va alors permettre comme il a été suggéré dans Sec. 2.2.2 d'implanter des calculs complexes (comme interpoler un visage inconnu depuis deux visages connus) simplement par la configuration spatiale de l'activité sur des surfaces (les cartes topologiques) représentant des gammes de stimulus complexes auto-organisés comme dans "l'espace de faces" observé dans IT.
Si l'on considère une assemblée de neurones, un tel codage nécessite une propagation de l'information entre ces neurones et donc d'implanter des *inter-*

*L'adjectif *épars* a été préféré ici à *parcimonieux* pour traduire l'anglais *sparse*.

actions latérales. De telles interactions semblent en fait jouer un rôle majeur dans le SNC puisque leur inactivation peut par exemple annihiler la propriété d'invariance au contraste de la courbe de sensibilité des neurones de V1 à l'orientation [Troyer et al., 1998] et donc à des mécanismes réduisant la redondance des réponses dans le cortex. Aussi, il a été montré que ces mécanismes jouent un rôle pour des neurones sélectifs à une orientation dans l'annihilation de la réponse à une direction perpendiculaire. Les interactions latérales rapides peuvent donc jouer un rôle dans une représentation robuste et rapide de l'information visuelle en vue de son traitement rapide.

Représentations sur-complète

À l'image d'une traduction d'un texte en plusieurs niveaux d'écriture (un texte technique par rapport à un texte de vulgarisation), la représentation sera d'autant plus éparse que le dictionnaire utilisé est adapté et riche. On pourra implanter un codage épars en utilisant une *représentation sur-complète*, c'est-à-dire en utilisant un dictionnaire dont la dimension sera de plusieurs magnitudes la dimension de l'espace a représenter. Typiquement, cette représentation sera reliée aux causes physiques du signal à représenter et on désirera notamment aux différents niveaux du traitement déterminer les causes indépendantes à l'origine du signal. Cette approche se rapproche alors d'un cadre probabiliste qui considère la probabilité conditionnelle de générer un signal en fonction des statistiques des signaux, qui sont considérés comme une réalisation d'un modèle génératif.
Dans la suite, on se ramènera à une représentation linéaire, et dans le cadre d'un modèle génératif linéaire, une implantation d'un codage épars consiste à considérer des filtres ϕ_i pour $i \in \mathcal{D}$, où $\mathcal{D}$ est un dictionnaire sur-complet et de chercher une approximation $\tilde{I}$ de l'image I de la forme

$$\tilde{I} = \sum a_i \phi_i \tag{4.1.1}$$

telle que les coefficients a_i sont nuls pour une grande majorité des filtres [Olshausen et Field, 1998]. Dans le cadre du codage dynamique que nous avons défini, nous allons plus particulièrement chercher une représentation

telle que $\tilde{I}(0) := 0$ et

$$\tilde{I}(t+1) = \tilde{I}(t) + a_{i^t}\phi_{i^t} \tag{4.1.2}$$

avec des coefficients a_{i^t} décroissant le plus rapidement vers zéro[†].
Pour résoudre le choix des filtres et des coefficients, il est alors nécessaire de définir un critère qui prend généralement la forme d'une distance $\|I - \tilde{I}\|$ entre l'image et sa reconstruction. Dans le problème à dimension finie qui nous intéresse, cela revient à choisir une sous-base comme un sous-ensemble du dictionnaire et qui représente optimalement l'image au sens de cette distance. Or il a été montré que ce problème conduit à une explosion combinatoire avec le nombre d'éléments du dictionnaire, et qu'en fait ce problème est *NP-complet*[‡] [Mallat, 1998, p.409]. Des algorithmes comme le *Basis Pursuit* [Chen et Donoho, 1994] peuvent toutefois trouver la base optimale pour représenter l'image, mais son implantation est prohibitive calculatoirement pour des images et n'a aucun support biologique.

4.1.2 Codage par poursuite

Nous allons alors montrer comment construire un modèle de codage surcomplet en définissant des interactions latérales adaptées selon un algorithme de poursuite.

À la recherche de la construction d'interactions latérales

En effet, comme nous voulons caractériser la "signature" de l'image, une première approche est de répétitivement estimer le point correspondant à une activité maximale puis d'inhiber les points voisins (et dont l'activité des filtres corrélés va aussi être forte) pour éviter que le processus se répète aux même points. Cette méthode est par exemple utilisée par Itti et al. [1998], sous le terme d'*inhibition de retour* et permet après avoir choisi un point

[†]En pratique on mesurera le caractère épars par la norme L-p, $\|a\|_p = [\sum_{1\leq i\leq t} |a_i|^p]^{\frac{1}{p}}$ pour p proche de 0.

[‡]C'est à dire qu'il n'existe pas d'algorithme à temps polynomial qui puisse donner la solution optimale.

saillant de l'image d'inhiber son voisinage sur un certain rayon et de passer à une nouvelle partie de l'image. Toutefois, cette inhibition est la même pour tous les points maximaux et la suite de filtres choisis ne permet pas de reconstruire l'image. Si par exemple deux points proches sont également saillants, cette inhibition cachera un des deux points et effacera indifféremment l'activité au voisinage du maximum.
Une autre approche consiste à utiliser les régularités de l'image pour en retirer les redondances. Elle a été formalisée par Barlow et Foldiak [1989] et consiste à utiliser les statistiques des images naturelles pour obtenir des réponses décorrélées, c'est-à-dire moins dépendantes (voir Sec. 3.2.3). Toutefois, cette méthode s'approche plus d'une analyse en composantes principales alors que les signaux visuels possèdent des corrélations d'ordre supérieur qui ne sont pas extraites par ce processus.
Enfin, une autre méthode consiste à apprendre la "forme" de l'inhibition [Li, 1998] et d'utiliser un processus répétitif qui va choisir les filtres les plus adaptés puis d'inhiber latéralement les voisins suivant cette forme avant de recommencer le processus. Plus précisément, on peut choisir d'inhiber latéralement l'activité suivant la corrélation des filtres entre eux : il est en effet plus logique d'inhiber plus fortement les filtres fortement corrélés au filtre choisi. C'est cette méthode qui nous a conduit à étudier les interactions latérales comme le support d'un algorithme de poursuite.

Poursuite gourmande

En effet, ces mécanismes se rapprochent d'un algorithme de *poursuite gourmande* qui est une technique introduite en statistique pour calculer un modèle génératif de densités de probabilités. On va estimer les filtres et coefficients d'Éq. 4.1.2 en utilisant une stratégie récursive "gourmande" (ou "gloutonne") : (1) on choisit le point correspondant à l'activité la plus forte, puis (2) on inhibe l'image suivant le filtre qui a été sélectionné, en soustrayant à l'image la projection orthogonale du filtre. Cet algorithme va donc introduire un processus récursif générant d'un coté une suite de résidus de l'image à décomposer et d'un autre une estimation linéaire par les éléments

du dictionnaire.
Formellement, on va donc dans un premier temps calculer les activités C_i pour tous les filtres en utilisant le produit scalaire défini par un espace de Hilbert $\mathcal{H}$, ici le produit scalaire usuel :

$$\begin{aligned} C_i &:= < I, \phi_i > & (4.1.3) \\ &= \sum\nolimits_{\vec{l} \in \mathcal{R}_i} I(\vec{l}).\phi_i(\vec{l}) & (4.1.4) \end{aligned}$$

Ensuite, on va choisir le filtre i^0 correspondant à l'activité maximale, ce que l'on note :

$$i^0 = \text{ArgMax}_i(|C_i|) \tag{4.1.5}$$

On retire alors de l'image la projection de ce filtre (c'est en effet la solution optimale pour minimiser la distance dans l'espace $\mathcal{H}$) pour obtenir une nouvelle image I^1 (le résidu).

$$\begin{aligned} I^1 &= I - \frac{< I^0, \phi_{i^0} >}{\|\phi_{i^0}\|^2} \phi_{i^0} & (4.1.6) \\ &= I - C^0_{i^0} \frac{\phi_{i^0}}{{N_{i^0}}^2} & (4.1.7) \end{aligned}$$

Aussi, on peut calculer l'effet de l'inhibition sur l'activité par :

$$\begin{aligned} C^1_i &= < I^1, \phi_i > & (4.1.8) \\ &= C_i - C_{i^0} \frac{< \phi_{i^0}, \phi_i >}{{N_{i^0}}^2} & (4.1.9) \end{aligned}$$

Cette dernière équation montre alors qu'on peut directement retirer de l'activité la corrélation d'un filtre par rapport au filtre sélectionné et continuer la procédure récursive sans utiliser Éq. 4.1.5. Si on initialise l'image et l'activité à $I^0 := I$ et $C^0_i = C_i$, alors il vient en répétant les procédures précédentes pour $t \geq 0$:

$$\begin{cases} i^t = \text{ArgMax}_{i \in \mathcal{D}}(|C^t_i|) \\ I^{t+1} = I^t_i - C^t_{i^t}.\frac{\phi_{i^t}}{{N_{i}}^{t^2}} \\ C^{t+1}_i =< I^{t+1}, \phi_i > \end{cases} \tag{4.1.10}$$

ou encore de façon équivalente en calculant directement l'activité résiduelle :

$$\begin{cases} i^t = \text{ArgMax}_{i\in\mathcal{D}}(|C_i^t|) \\ C_i^{t+1} = C_i^t - C_{i^t}^t.\frac{<\phi_{i^t},\phi_i>}{{N_{i^t}}^2} \end{cases} \tag{4.1.11}$$

Enfin, on pourra alors reconstruire simplement le signal comme la somme des filtres que nous avons sélectionnés :

$$\tilde{I}^{t+1} = \tilde{I}^t + C_{i^t}^t.\frac{\phi_{i^t}}{{N_{i^t}}^2} \tag{4.1.12}$$

soit

$$\tilde{I}^T = \sum_{1\geq t\geq T} C_{i^t}^t.\frac{\phi_{i^t}}{{N_{i^t}}^2} \tag{4.1.13}$$

et qui correspond bien à notre but dans Éq. 4.1.1 et Éq. 4.1.2. On justifie le terme "gourmand" par le fait qu'on choisit à un instant donné le meilleur filtre sans se soucier des conséquences à long terme.
L'algorithme que nous avons construit est alors exactement une *Poursuite de Concordance* ou *Matching Pursuit* (MP) [Mallat, 1998, pp.412–9] pour des filtres normalisés ($N_i = 1$ pour tout i). Cet algorithme est particulièrement adapté à des approximations adaptatives non-linéaires (il peut être étendu à d'autres produits scalaires arbitraires) et, malgré sa simplicité, montre des comportements complexes et chaotiques [Davis, 1994].

Propriétés de la poursuite gourmande

Tout d'abord, on remarque qu'à chaque pas, l'activité correspondante d'un filtre est bien annulée :

$$C_{i^t}^{t+1} = C_{i^t}^t - C_{i^t}^t.\frac{<\phi_{i^t},\phi_{i^t}>}{{N_{i^t}}^2} = 0 \tag{4.1.14}$$

Aussi, comme on soustrait la projection orthogonale, une propriété de la poursuite est alors que par le théorème de Pythagore

$$\|\tilde{I}^t\|^2 = \|\tilde{I}^{t+1}\|^2 + |C_{i^t}^t|^2.\|\frac{\phi_{i^t}}{{N_{i^t}}^2}\|^2 \tag{4.1.15}$$

soit

$$\|\tilde{I}\|^2 = \sum_{1\geq t\geq T} |\frac{C_{i^t}^t}{{N_{i^t}}^2}|^2 + \|\tilde{I}^{t+1}\|^2 \tag{4.1.16}$$

qui correspond à une formule de conservation de l'énergie. Grâce à cette relation, Mallat et Zhang [1993] ont prouvé que cet algorithme convergeait pour un dictionnaire au moins complet dans l'espace $\mathcal{H}$, de telle sorte que

$$\lim_t \|I - \tilde{I}^t\|^2 = 0 \tag{4.1.17}$$

De plus, cette décroissance est exponentielle, et sa vitesse dépend du caractère sur-complet et adapté du dictionnaire. De façon plus générale, la convergence est assurée sur l'espace généré par les filtres de la base.
Un premier avantage de la poursuite gourmande est qu'on va pouvoir contrôler le caractère épars de la représentation en modifiant la richesse du dictionnaire. Les deux extrêmes vont alors d'un coté correspondre à une représentation dense par un dictionnaire orthogonal : puisque la corrélation entre les filtres est nulle on n'inhibera seulement l'activité du filtre choisi et on se ramènera donc à l'algorithme de Van Rullen et Thorpe [2001]. De l'autre coté, si l'on choisit un dictionnaire correspondant à toutes les images possibles, l'algorithme va converger exactement en une itération. Ces deux stratégies extrêmes (représentation distribuée ou locale) vont correspondre à des quantités d'information basses ou très élevées de la suite des filtres et coefficients choisis. On adoptera ici avec un codage épars une stratégie intermédiaire (voir [Foldiak et Young, 1995] pour une revue des avantages de cette stratégie) dont il conviendra de quantifier la qualité.
Toutefois, cet algorithme peut fournir des solutions optimales localement. En effet, comme l'algorithme est gourmand et qu'il dépend de la séquence de filtres sélectionnés, un mauvais choix pourra se propager aux itérations suivantes. Une amélioration consiste en premier lieu à orthogonaliser a posteriori la base des filtres qu'on a choisis [Pati et al., 1993]. On peut alors montrer que cette stratégie revient à chercher des solutions dans l'espace conjugué à

celui constitué par la base déjà sélectionnée*. Toutefois, cette stratégie est computationnellement coûteuse et s'écarte d'un modèle biologique.

4.1.3 Code impulsionnel épars

Comme l'avait pressenti Mallat et Zhang [1993], cet algorithme peut alors être interprété dans la structure d'un réseau neural. Nous allons montrer qu'il s'applique particulièrement à un code impulsionnel et permet grâce à la définition d'interactions latérales de construire un code impulsionnel épars.

Codage impulsionnel

En effet, on peut interpréter l'algorithme de codage par poursuite gourmande dans le cadre d'une couche de neurones à impulsions. Tout d'abord comme pour Sec. 3.2.1, l'activité initiale C_i^0 des filtres i peut être calculée grâce au modèle linéaire que nous avons présenté. Aussi comme nous l'avons vu, le premier neurone à émettre un PA sera celui dont l'activité sera la plus forte, et donc le premier PA correspond au premier filtre sélectionné i^0. Selon Éq. 4.1.5, cela correspondra à :

$$i^0 = \text{ArgMax}_{i \in \mathcal{D}}(|C_i^0|) \tag{4.1.18}$$

On définit alors une couche d'interactions latérales comme des interactions très rapides* et définies par la corrélation entre les filtres. Une règle de type hebbien semble alors parfaitement adaptée pour calculer ces poids puisqu'elle correspond exactement à la corrélation $< \phi_{i_1}, \phi_{i_2} >$ entre les filtres de deux neurones i_1 et i_2. Dans ce modèle théorique, nous inhiberons alors immédiatement l'activité des filtres corrélés avec le filtre sélectionné.

$$C_i^1 = C_i^0 - C_{i^0}^O . \frac{< \phi_{i^0}, \phi_i >}{{N_{i^0}}^2} \tag{4.1.19}$$

La propagation peut alors être poursuivie pour les itérations suivantes en

*La démonstration utilise le fait qu'on se place à chaque pas de temps de l'algorithme dans le sous-espace orthogonal à celui des filtres déjà choisis.

*par rapport à la constante de temps membranaires τ_m

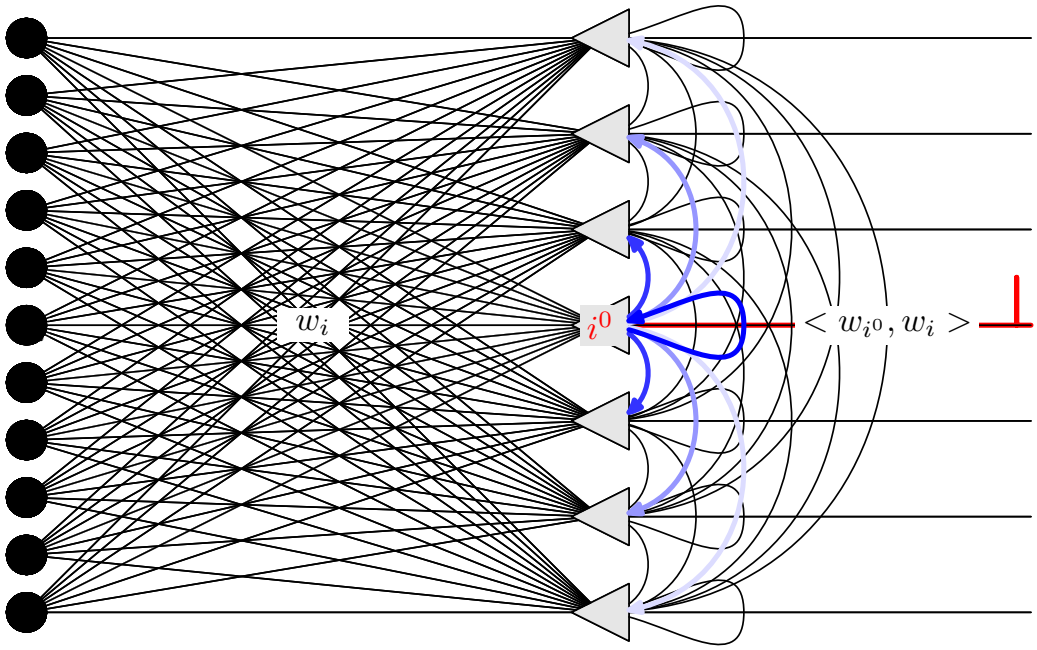

FIG. 4.1.2 – *Principe du codage impulsionnel par poursuite gourmande.* On a représenté un réseau de neurones i partageant des entrées similaires (points noirs) grâce à des poids $\vec{w}_i$. Le principe du codage par poursuite gourmande revient à choisir un neurone i^0 (par exemple parce qu'il correspond à la plus forte activité), qui va émettre une impulsion puis à soustraire à l'activité des autres neurones une quantité proportionnelle à leur corrélation $< \vec{w}_{i^0}, \vec{w}_i >$. La procédure est alors répétée récursivement, générant ainsi une suite de PAs qui vont représenter l'entrée.

modifiant ainsi seulement l'activité des neurones.

$$\begin{cases} i^t = \text{ArgMax}_{i \in \mathcal{D}}(|C_i^t|) \\ C_i^{t+1} = C_i^t - C_{i^t}^t . \frac{<\phi_{i^t}, \phi_i>}{{N_{i^t}}^2} \end{cases} \tag{4.1.20}$$

Cette stratégie est à la fois computationnellement plus efficace (au lieu de recalculer tous les produits scalaires, on ne doit que mémoriser initialement ces corrélations et effectuer des additions) mais aussi biologiquement plus réaliste.
On pourra aussi imaginer que la poursuite soit seulement α-*gourmande*, c'est-à-dire que chaque impulsion corresponde à une inhibition telle que

$$C_i^{t+1} = C_i^t - \alpha . C_{i^t}^t . \frac{< \phi_{i^t}, \phi_i >}{{N_{i^t}}^2} \tag{4.1.21}$$

L'information impulsionnelle pour $\alpha \in [0, 1[$ sera alors plus progressive (le modèle présente alors des trains d'impulsions similaires à la biologie, voir Fig. 4.1.3). D'un autre côté, un modèle avec $\alpha > 1$ présentera —comme pour un système différentiel du second ordre— une relaxation plus rapide mais en contrepartie plus instable. Nous nous limiterons dans la suite au cas correspondant à la poursuite gourmande classique ($\alpha = 1$).
Cet algorithme présente donc un modèle simple de codage non-linéaire impulsionnel. Il permet de prévoir des comportements neuraux biologiques qui ne correspondent pas à un modèle linéaire comme par exemple la réponse de neurones de V1 à des orientations. En effet, si l'on observe qu'un seul neurone dans notre modèle (comme ce qui est souvent une limite de la technologie d'enregistrement neurophysiologiques), l'activité va dépendre directement de l'activité des neurones corrélés. Son comportement va être influencé non-linéairement par l'ensemble des sensibilités de ces neurones, un comportement qui se rapproche des observations biologiques par rapport au modèle linéaire de Rodieck [1965].

Décodage impulsionnel

De la même manière, pour étudier la qualité d'un décodage impulsionnel, nous allons imaginer que l'on veuille reconstruire l'image même si —répétons-

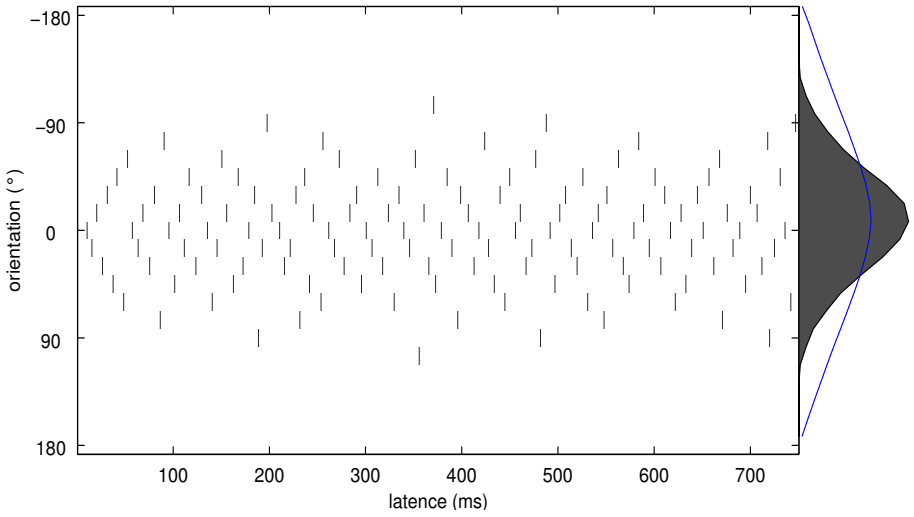

FIG. 4.1.3 – *Codage impulsionnel par poursuite α-gourmande.* On a simulé ici un réseau de 25 neurones LIF sélectifs à l'orientation (filtres de Gabor) auxquels on présente à l'instant $t = 0$ la direction préférée. En plus du mécanisme impulsionnel et de la remise à zéro, on a simulé une poursuite α-gourmande, pour $\alpha = 50\%$. La réponse fréquentielle sur 3 s est présentée à droite (courbe pleine) et est comparée avec la réponse linéaire. Cet exemple simple montre que la poursuite gourmande resserre donc la courbe de sensibilité des neurones (et on contrôle ce comportement par le paramètre α), et accentue l'expression temporelle d'un code impulsionnel, un comportement non linéaire qui est observé dans V1. Pour le cas $\alpha = 1$, le neurone correspondant à la direction préféré serait le seul à répondre. On remarque la similarité de la réponse impulsionnelle avec les résultats de Celebrini et al. [1993].

le— ce n'est pas réaliste biologiquement. Il suffira alors de considérer la même architecture que pour Sec. 3.2.2 et écrire la reconstruction grâce à la relation dans Éq. 4.1.12. C'est à dire qu'à chaque impulsion arrivant au récepteur, on pourra écrire

$$\tilde{I}^{t+1} = \tilde{I}^{t} + m^{t}.\phi_{i^t} \tag{4.1.22}$$

où m^t évalue la valeur du coefficient de rang t.
De manière similaire à Sec. 3.2.2, nous avons étudié la régularité de la décroissance de la valeur absolue des coefficients en fonction du rang de décharge. Mais on peut améliorer le code impulsionnel en imaginant alors que cette fonction de modulation est apprise aux niveaux du décodage mais aussi du codage, et ainsi profiter du caractère adaptatif de la poursuite gourmande. Ainsi, comme dans Sec. 3.2.2, on va calculer la table de correspondance par par exemple

$$m^{(n+1)}(t) = m^{(n)}(t) + \mu^{(n)}.(|C^t_{i^t}| - m^{(n)}(t)) \tag{4.1.23}$$

où n correspond au pas d'apprentissage de la table de correspondance. Toutefois, on va utiliser pour l'inhibition la valeur correspondant à celle qui est attendue au décodage. On aura donc une nouvelle expression de l'algorithme sous la forme

$$\begin{cases} i^t = \mathrm{ArgMax}_{i\in\mathcal{D}}(|C^t_i|) \\ C^{t+1}_i = C^t_i - m^{(n)}(t).<\phi_{i^t}, \phi_i> \end{cases} \tag{4.1.24}$$

Ces expressions définissent alors une stratégie complète de codage et décodage impulsionnel avec une représentation sur-complète et éparse en utilisant des interactions latérales.

4.1.4 Application au codage rétinien

Une fois présenté le modèle, nous allons pouvoir le comparer à celui que nous avions présenté à la Sec. 3.2 et présenter des extensions qui répondent aux problèmes que nous avions évoqué pour proposer un modèle impulsionnel

de codage rétinien.

Comparaison sur une grille dyadique

Tout d'abord, à des fins de comparaison, nous avons confronté les résultats du modèle par ondelette par rapport au modèle par poursuite gourmande sur une grille dyadique avec les mêmes filtres. Comme ces filtres sont quasiment orthogonaux, ces algorithmes vont être très proches. En particulier, si l'on choisit une base orthogonale alors les interactions latérales que nous avons définies se réduisent seulement à une inhibition du neurone qui a tiré.
Toutefois, si l'on considère maintenant le codage impulsionnel dans son ensemble, on remarque que la poursuite gourmande, par son caractère adaptatif va garantir une meilleure reconstruction de l'image (voir Fig. 4.1.4). Au niveau computationnel, l'implantation d'un tel code sur un calculateur classique requiert une grande quantité de mémoire pour mémoriser les filtres et toutes leurs corrélations. Pour une grille générée par translation de N échelles, on calcule les noyaux de corrélation par des convolutions entre chaque échelle, et donc par symétrie $N.(N+1)/2$ filtres. Ensuite les calculs consistent à chaque pas de temps en une recherche de maximum et une addition : cet algorithme est donc particulièrement adapté à un calculateur distribué. En particulier, il est adapté à des architectures neuromorphiques basées sur des sous-systèmes élémentaires fonctionnant suivant un principe compétitif comme le "Gagant Prend Tout" (*Winner Take-All*) [Indiveri, 2001].

Extensions à des modèles réalistes

De plus, nous allons pouvoir élargir le champ des modèles rétiniens grâce au caractère souple de cet algorithme.
Tout d'abord, on peut imaginer implanter en plus des filtres rétiniens que nous avons décrit des filtres répondant particulièrement à un bruit parasite, c'est-à-dire à des éléments indésirables et connus du signal comme par exemple des discontinuités fortes liées à des rayures pour des documents numérisés. En effet, il peut être utile d'éliminer des signaux aberrants en intégrant des filtres sélectifs au bruit à l'algorithme que nous avons présenté

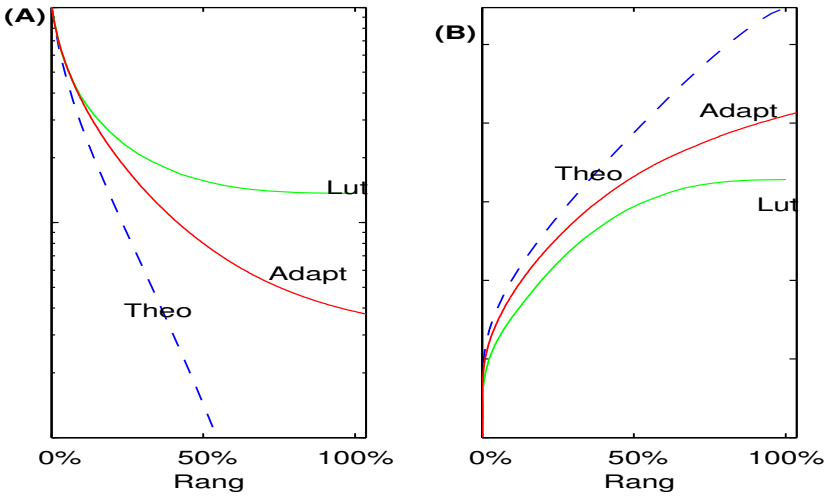

FIG. 4.1.4 – *Comparaison ondelettes / poursuite gourmande sur une grille dyadique.* On a comparé la mesure progressive en fonction du rang de **(A)** l'Erreur Quadratique moyenne (EQM) et **(B)** l'information mutuelle (IM) pour les différentes stratégies de code impulsionnel. **(Lut)** Reconstruction par les coefficients des ondelettes. **(Adapt)** Reconstruction par le code impulsionnel. Le caractère adaptatif du code par poursuite gourmande, mais aussi sa structure qui va progressivement éliminer les dépendances entre les filtres, induit une meilleure transmission de l'information, proche de **(Théo)** la reconstruction par les coefficients.

mais en les éliminant de la propagation. Ainsi, on ne propagera que les interactions qui inhibent les effets du bruit sur l'activité et sans propager l'impulsion associée au bruit.
Ensuite, on peut aussi modifier le critère de choix dans l'algorithme et donc l'ordre du choix des filtres choisis. Ainsi, on pourra par exemple normaliser les filtres de telle façon à ce que certains filtres dans l'espace temps / échelle soient privilégiés par rapport à d'autres. D'après ce que nous avons vu, le but sera par exemple de maximiser l'entropie de la sortie et dans ce cas, on pourra utiliser les méthodes introduites par Atick [1992] que nous avons exposées à la Sec. 3.2.3. Cette méthode révèle alors des résultats similaires avec le gain introduit par le caractère adapté de la méthode. Il se rapproche dans la pratique de techniques utilisées pour améliorer les performances de la poursuite de concordance en privilégiant les hautes fréquences [Jaggi et al., 1995].
Enfin, une implantation plus réaliste de la rétine et en accord avec les besoins de réduction de dimension pourrait limiter le nombre de filtres, en contraignant par exemple la taille suivant l'excentricité depuis le centre optique. Typiquement, ce nouveau pavage irrégulier est le pavage *log-polaire* de la surface de la rétine et peut s'adapter à des contraintes (comme le scotome) ou des lésions. L'algorithme présenté reste toujours valable dans ce cas, les résultats de Mallat et Zhang [1993] indiquant que la représentation sera parfaite dans l'espace généré par les filtres utilisés.

Pavage en quinconces

Dans le cadre d'un modèle de la rétine pour des images classiques sur une grille rectangulaire, un *pavage en quinconces* se trouve être un bon compromis entre la représentation sur-complète et la contrainte de réduction de dimension. Ce pavage se définit simplement par sa construction depuis la grille d'échelle 1 comme des transformations combinant une dilatation de $\sqrt{2}$ et d'une rotation de $\pi/4$, *i.e.* la grille des échelles impaires est la grille dyadique*. Par rapport à une grille dyadique où on divise le nombre de points

*De même, une grille multiéchelle adaptée à un maillage hexagonal pourra utiliser une une dilatation de $\sqrt{3}$ et d'une rotation de $\pi/3$.

par 4, le nombre de points est divisé par deux entre deux échelles.
On en déduit alors que la matrice de transformation des grilles d'une échelle à la suivante est simplement

$$\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

On en déduit alors bien que le nombre total de neurones sera supérieur au nombre de pixels d'un facteur multiplicatif de $\chi = \frac{1-(1/\sqrt{2})^{s\max}}{1-(1/\sqrt{2})^2}$ soit environ 2. Comme il existe une seule décomposition binaire d'un entier, il existe une manière simple de représenter l'activité sur une image de 2 couches (ou à double résolution en décalant la deuxième couche de $(1/2, 1/2)$) : une couche correspondant à la première échelle et les points de la deuxième correspondant un à un aux points des échelles suivantes, et qui sera adaptée à un codage compact. Pour cela, il faut translater la grille à l'échelle k générée par une dilatation de $2^{k/2}$ et d'une rotation de $k.\pi/4$ par exemple au centre d'un maillage non plein de la grille précédente.
Enfin, pour terminer la construction de cette transformation, on peut alors facilement créer une pyramide en créant une transformée récursive par une fonction d'échelle adaptant le signal à une grille plus grossière puis par une fonction d'ondelette (voir par exemple [Mallat, 1998, p.153]). Les interactions entre les filtres sont de la même manière définis récursivement. Cette définition est donc moins restrictive que la définition de la pyramide laplacienne, mais on choisira des filtres semblables adaptés à une bonne représentation de l'information dans cette pyramide en quinconces. En effet, l'adaptation de ces filtres accélérera d'autant la convergence de l'algorithme.

4.2 Modèle neuronal de représentation corticale sur-complète dans l'aire visuelle primaire (V1)

La section précédente a donc montré que l'on pouvait définir des modèles élaborés de code rétinien impulsionnel en adaptant un algorithme de pour-

suite gourmande pour modéliser les interactions latérales. Toutefois ces résultats doivent être compris à la lumière des contraintes imposées par la nature de la rétine et du goulot d'étranglement informationnel du nerf optique. Nous allons voir que cet algorithme prend alors toute sa dimension dans un modèle de représentation corticale que nous allons limiter dans un premier temps à l'aire visuelle primaire, V1.

4.2.1 Le modèle de Hubel et Wiesel

Une représentation des contours de l'image

L'aire visuelle primaire est une partie du cortex située dans la région occipitale du cerveau. Comme nous l'avons décrite dans la Sec. 2.2.2, cette aire joue un rôle majeur tant par sa position que par sa taille relative. Par rapport au million de fibres quittant la rétine, cette aire comprend chez l'homme près de 300 millions de neurones. Cette aire va donc être essentielle pour former les représentations qui vont ensuite se propager vers les aires secondaires (V2, V4) et supérieures (comme IT). Par rapport à la rétine, il n'existe plus de contrainte aussi forte sur le nombre de neurones et nous pourrons utiliser un dictionnaire sur-complet.

Depuis les travaux pionniers de Hubel et Wiesel [1959], on sait que de nombreuses cellules de cette aire sont sélectifs aux orientations des contours représentés par la rétine [Ben-Yishai et al., 1995]. Dans le cadre du traitement rapide, nous nous intéresserons donc particulièrement à ces cellules, dites simples, qui vont répondre à une orientation préférentielle. Nous avons déjà rencontré de telles cellules dans des modèles où elles étaient isolées (voir Sec. 3.3.3) ou encore une application où de telles cellules émergeait d'un apprentissage hebbien dans le cadre du codage par rang, et nous allons maintenant essayer de définir un modèle de V1 comme des couches de neurones sensibles à différentes orientations et à différentes échelles.

Pour cela, nous allons modéliser les cellules simples par des filtres linéaires comme des fonctions de Gabor [1946] définies comme une sinusoïde orientée (suivant une direction $\vec{\alpha}$ et une fréquence ν), de phase ϕ et modulée par une

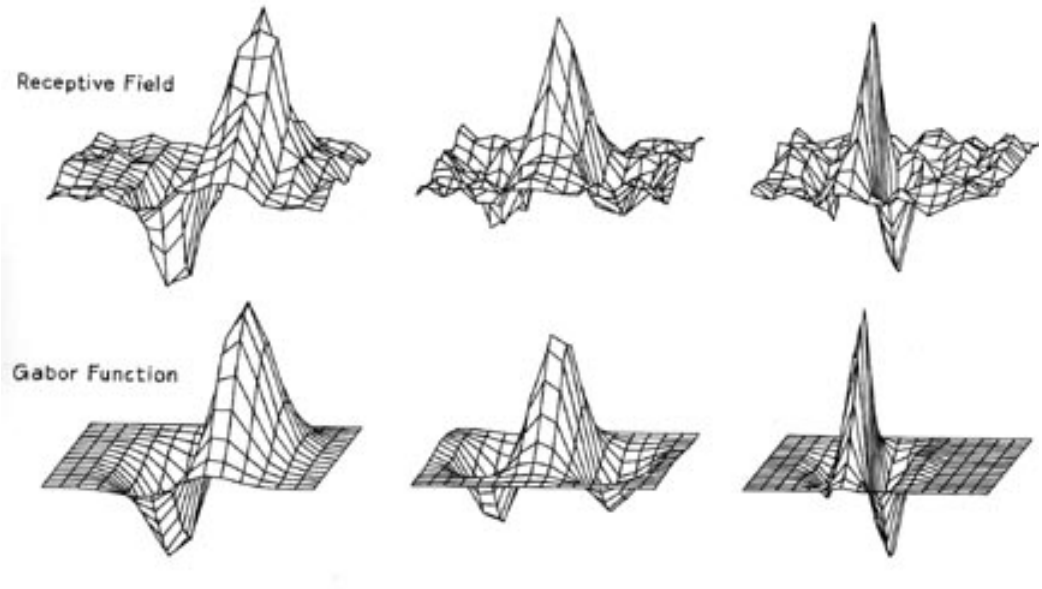

FIG. 4.2.1 – *Champs récepteurs des neurones de V1 chez le chat.* **(Haut)** La réponse impulsionnelle des neurones de V1 peut être reconstruite par des méthodes neuro-physiologiques. **(Bas)** Elle fait apparaître des champs récepteurs similaires à des fonctions de Gabor de phases, orientations et fréquences différentes. (d'après [Daugman et Downing, 1995]).

gaussienne de variance σ (voir [Mallat, 1998, p. 160])

$$\psi(\vec{r}) = \sin(\nu.(\vec{r}.\vec{\alpha}) + \phi).\frac{1}{\sigma\sqrt{\pi}}.\exp(-\frac{\|\vec{r}\|^2}{2.\sigma^2}) \tag{4.2.1}$$

comme il a été observé dans l'aire V1 de nombreux mammifères [Daugman et Downing, 1995] (voir Fig. 4.2.1).
Dans un premier temps, en accord avec les résultats neuro-physiologiques et pour limiter la charge computationnelle, nous utiliserons une grille en quinconce avec des filtres de Gabor symétriques ou antisymétriques (*i.e.* $\phi = 0$ ou $\phi = \pi/2$) de fréquences contraintes par la variance de la gaussienne sur 15 orientations différentes réparties uniformément sur $[0, 2\pi[$. Cette représentation est donc sur-complète d'une magnitude $2 \times 2 \times 15 = 60$. De la même manière que pour le code rétinien, nous avons choisi d'utiliser une optimisation statistique analogue des contrastes. En effet, si l'on suppose une invariance par rotation de la loi de distribution de l'énergie fréquentielle, on pourra optimiser l'entropie de la représentation en ajustant le gain des différentes orientations seulement en fonction de leurs tailles.

Code impulsionnel de V1 par poursuite gourmande

Cette architecture a été utilisée sur des images en niveaux de gris et nous avons étudié de la même façon la reconstruction par cet algorithme pour pouvoir en étudier la qualité et le comparer au codage rétinien. Le décodage par les coefficients calculés montre que comme attendu, la reconstruction est beaucoup plus rapide (voir Fig. 4.2.2) et donc que la représentation est plus éparse. En effet, on voit alors que quelques centaines d'impulsions sont suffisantes pour représenter l'image. En effet on observe sur l'histogramme de réponse des neurones que les réponses sont très sélectives (ces distributions ont une *kurtosis* élevée) et donc que chaque impulsion, soit la réponse d'un filtre, comportera d'autant plus d'information.
Comme précédemment, on peut construire un code impulsionnel complet en étudiant la régularité de la valeur absolue des coefficients en fonction de ler rang de tir. L'architecture utilise un nombre limité de filtres pouvant simuler

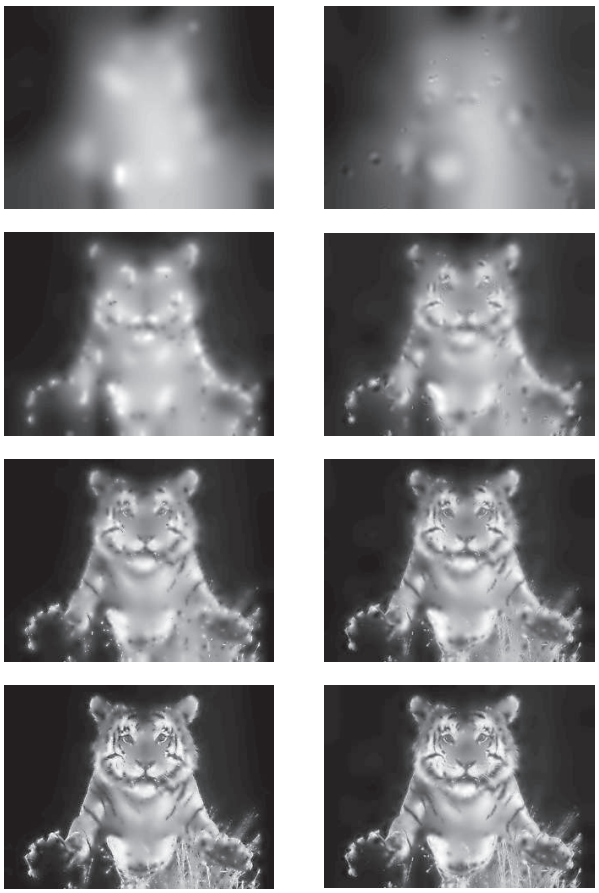

FIG. 4.2.2 – *Reconstruction de l'image par une poursuite gourmande soit dans la rétine soit dans V1.* On a comparé pour une image naturelle (**Gauche**) le codage rétinien et (**Droite**) le codage par des filtres sélectifs à des orientations ressemblant aux cellules simples rencontrées dans V1. On a montré la reconstruction de l'image après resp. 100, 750, 3000 et 9000 PAs. L'image originale est reconnue après quelques centaines d'impulsions seulement, les coefficients décroissant en valeur absolue rapidement vers 0. Les contours sont propagés en avant dans V1 et l'information est propagée plus rapidement.

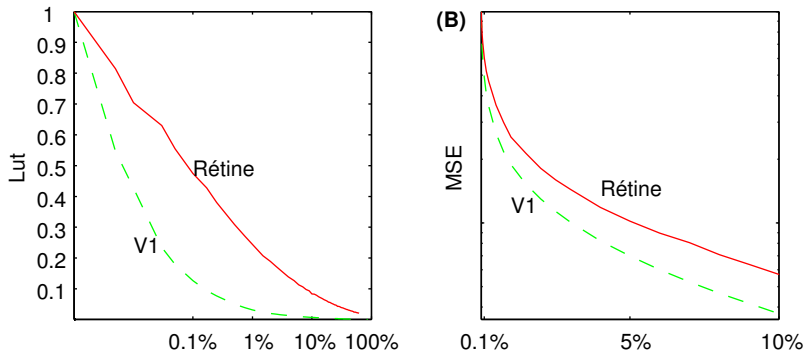

FIG. 4.2.3 – *Code impulsionnel dans V1.* **(A)** On a calculé de manière récursive la fonction de modulation donnant la valeur absolue d'un coefficient en fonction de son rang. **(B)** La reconstruction de l'image par cet algorithme prouvent qu'on a ainsi défini un code visuel de V1 par des filtres de Gabor qui conduisent à une représentation impulsionnelle éparse.

---

les différentes orientations, et nous avons en particulier utilisé l'algorithme de *steerable pyramid* [Freeman et al., 1989]. Comme ces calculs restent particulièrement lourds sur un calculateur classique, on a limité l'apprentissage et la reconstruction à 20 images. Toutefois, les résultats restent similaires à la rétine qualitativement, même si la convergence est beaucoup plus rapide.

Introduction de connaissance a priori : coopération locale

On remarque en pratique qu'un neurone de V1 a d'autant plus de chances de tirer qu'un neurone voisin a déjà tiré, et cela malgré l'inhibition des neurones corrélés. De manière plus précise, on remarque que dans les images naturelles, les contours orientés sont le plus souvent continus et —qu'en dehors d'une texture aléatoire— il existe donc des corrélations d'ordre supérieur liées à la continuité des lignes de contours et par conséquent entre des neurones voisins sensibles à des orientations voisines. Il est alors désirable dans la cadre du traitement rapide de prendre en compte cette connaissance a priori de la structure des images pour propager en avant les contours continus des objets. Cette démarche se rapproche alors du principe d'hystérésis dans l'algorithme de détections des contours de Canny [1986].

Du côté neurophysiologique, des études récentes se rapprochent de cette stratégie. En effet, (Berry, et al. 1999) ont montré que dès la rétine, les cellules ganglionnaires pouvaient être sensibles à des mouvements qui pouvaient prédire la prochaine position du stimulus visuel (la corrélation en jeu est alors directement reliée à la continuité des mouvements physiques). Plus particulièrement, le seuil de sélectivité des neurones de V1 peut être modifié en fonction de l'activité du voisinage [Li, 1998]. Nous allons donc proposer un modèle qui va baisser le seuil de tir (ou réflectivement, qui va augmenter la norme relative des filtres dans le critère de choix comme évoqué précédemment) des neurones susceptibles d'être reliés par cette corrélation d'ordre supérieur. Ce profil va être appris sur des images synthétiques présentant des bords continus, comme des images de cercles de centres et rayons aléatoires.

Nous allons alors appliquer ce paradigme à une image de synthèse aléatoire mais, par réalisme, de même spectre d'énergie que les images naturelles. Les intensités sur un disque sont inversées de telle sorte que subjectivement, le disque apparaît comme l'image d'un objet avec des textures différentes (voir Fig. 4.2.4-Gauche). L'algorithme de codage que nous avons présenté va alors propager les parties de l'image en les choisissant de manière quasi-aléatoire (voir Fig. 4.2.4-Haut) mais en privilégiant d'abord les discontinuités. En introduisant une connaissance a priori (par exemple en favorisant les segments appartenant à de mêmes cercles [Sigman et al., 2001]), on peut déclencher de proche en proche une "avalanche" de propagation suivant le cercle et qui tend à le révéler alors en premier (voir Fig. 4.2.4-Bas) tout en préservant les propriétés de convergence de l'algorithme puisqu'on ne fait que modifier le seuil et donc l'ordre de tir. On en déduit alors que la représentation fera apparaître le cercle de manière transiente en propageant préférentiellement les singularités colinéaires. En se référant aux arguments de Bullier et Nowak [1995] et de Van Rullen et al. [2001], cet algorithme n'introduit pas à proprement dire de *feedback*, *i.e.* de retour en arrière de l'information et est donc toujours compatible avec une propagation en avant de l'information. Il permet de prendre en compte la connaissance a priori par une coopération locales des neurones ayant des sensibilités voisines et en mettant en avant

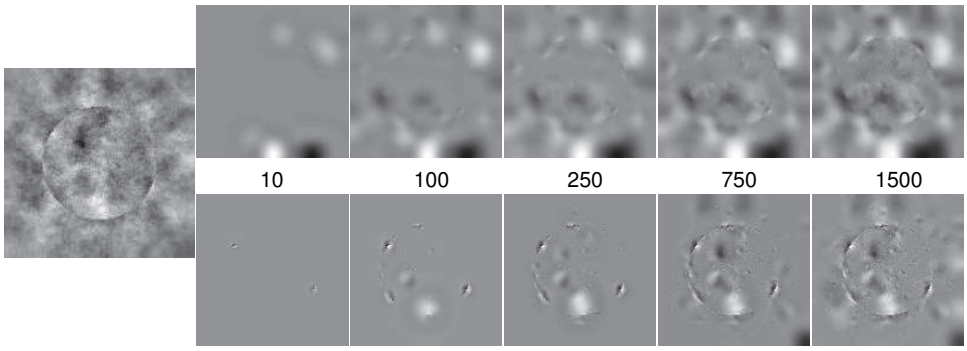

FIG. 4.2.4 – *Introduction de connaissance a priori.* **(Haut)** Une image de synthèse est construite en dessinant un disque avec une texture aléatoire sur un fond de même nature. **(Gauche)** Une poursuite gourmande détecte et propage les filtres selon un ordre quasi-aléatoire, mais en propageant les discontinuités en premier. **(Droite)** En introduisant une connaissance a priori, on favorise les bord colinéaires : la reconstruction est donc modifiée pour mettre en avant le cercle. Toutefois, la convergence de la propagation est préservée.

(dans le temps) les structures cohérentes de l'image. En plus d'améliorer la propagation du flux d'information, cette technique peut être utilisée pour optimiser la compression du signal à transmettre [Buccigrossi et Simoncelli, 1997].

4.2.2 Modèles d'adaptation

De manière similaire au modèle utilisant une transformée par ondelettes, ce modèle permet de convertir un signal lumineux en une vague d'impulsions. Celles-ci vont coder progressivement et dans le temps des valeurs analogiques et il est donc primordial de contrôler la compétition temporelle entre les neurones pour maximiser la transmission de l'information.

Adaptation neuronale

Les algorithmes que nous avons présenté sont basé sur un modèle génératif de l'image. On comprend alors l'importance des phénomènes adaptatifs qui vont transformer non-linéairement les activités pour que ce modèle soit va-

lable. En premier lieu, on choisira donc d'utiliser la luminosité des pixels pour pouvoir établir des contrastes par des filtres linéaires dans la rétine. De la même façon, on choisira de modifier la sensibilité des filtres aux différentes tailles comme décrit dans Sec. 3.2.3.
On remarque alors que si on imagine un système dynamique composé de plusieurs couches neurales, l'adaptation introduite dans une couche sera propagée aux couches suivantes. Ainsi pour un système composé de la rétine et de V1, si l'on optimise la norme des filtres dans la rétine, alors la reconstruction fera apparaître cette optimisation.
Ainsi dans l'aire visuelle primaire, certaines orientations peuvent être privilégiées suivant les statistiques des images naturelles. En particulier, les images produites dans des environnements artificiels (ville, laboratoire) présentent avant tout des horizontales et des verticales. Si on veut optimiser la taille de la représentation d'une image, on pourra donc normaliser en conséquence les filtres en fonction de leur orientation préférée.

Auto-organisation spatiale des filtres

Cet algorithme est particulièrement adapté à des systèmes physiques de calcul parallèle. Si dans un calculateur synchrone, on simule ce système en répliquant les configurations des neurones (paramètres, poids, connexions latérales), dans un système biologique des règles d'adaptation locales permettent de déterminer des jeux de paramètres locaux. Dans cette optique, il est alors intéressant d'étudier des règles d'auto-organisation spatiale de tels neurones et de voir comment elles peuvent s'adapter et coopérer avec le principe de codage que nous avons exposé.
Nous avons donc simulé un modèle d'aire visuelle primaire sur une grille rectangulaire de la taille d'images 128×128 comportant un filtre de Gabor (parmi 8 orientations différentes) et un seul sur chaque point de l'image. Dans cette simulation , on essaye d'optimiser la répartition spatiale —initialement aléatoire— de ces filtres à différentes polarités afin d'optimiser la représentation de l'image. D'un coté, comme à un point correspond une seule polarité, il faut que les différents filtres soient assez distribués pour couvrir localement

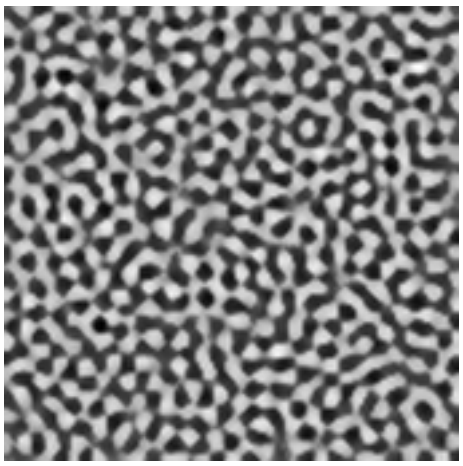

FIG. 4.2.5 – *Auto-organisation spatiale des polarités dans la rétine.* On a modélisé une couche de neurones sensibles à des contrastes (suivant des DDG Éq. 3.2.2) où chaque pixel correspond une polarité possible. La polarité d'un neurone évolue suivant un renforcement en présentant des images naturelles transformées par le code impulsionnel défini par la poursuite gourmande. Cette adaptation met en jeu une coopération locale et une compétition à plus longue distance de telles sorte que la simulation fait apparaître des bandes de Turing comme il est observé dans la rétine ou les aires corticales.

chaque point. D'un autre coté, les filtres coopèrent localement et on tendance à s'arranger par sensibilité voisines.
Ce processus d'adaptation va donc combiner une excitation à courte portée et une inhibition plus large qui sont caractéristiques des systèmes étudiés par Turing [1952]. Ainsi, une simulation sur des images naturelles révèle bien l'émergence de *bandes de Turing*, comme ce qui est observé dans de nombreux systèmes biologiques comme aussi l'aire visuelle primaire (voir Fig. 4.2.5).

Émergence de filtres et séparation aveugle de sources

Le principe de l'algorithme de poursuite de concordance peut se rapprocher des algorithmes de Quantification par Vecteurs. En effet, ce dernier algorithme est utilisé pour représenter un signal par des vecteurs "prototype" tirés d'un dictionnaire. Par exemple, on peut quantifier la couleur d'un pixel par les trois composantes rouge, vert et bleu : ces composantes seront

alors les vecteurs formant le dictionnaire des couleurs. Cet algorithme est appliqué notamment quand les données à représenter se présentent sous forme de grappes (ou *clusters*), les vecteurs de la base (ou centroïdes) formant le centre -ou noyau- de chaque grappe. Aussi, cet algorithme est associé à une règle d'apprentissage non-supervisé similaire à une stratégie de poursuite. En effet, l'*algorithme Généralisé de Lloyd*, en partant d'une base aléatoire de vecteurs, consiste à récursivement (1) choisir les données se rapprochant des centroïdes (2) rapprocher les centroïdes vers le centre de chaque grappe définie. L'algorithme est arrêté lorsque la distorsion entre les données et sa représentation par les centroïdes est assez faible.
On peut alors transposer cet algorithme à notre code impulsionnel. Nous allons d'abord choisir au hasard la base de départ (*i.e.* les filtres). Nous décomposerons alors grâce au code impulsionnel des images tirées aléatoirement dans une base de donnée d'images au moyen de ces filtres. Ceux-ci seront alors actualisés par une règle hebbienne vers la zone de l'image limitée à leur champ récepteur $\mathcal{R}_t$ qui les a excité et ce d'autant plus rapidement que l'activité est importante (vitesse réglée grâce au pas d'apprentissage γ). Formellement, au pas t :

$$\phi_{i^t} \leftarrow (1-\gamma).\phi_{i^t} + \gamma.\frac{{N_{i^0}}^2}{C_{i^t}}.I_{\mathcal{R}_t} \tag{4.2.2}$$

La poursuite est interrompue quand le résidu est assez faible et l'apprentissage est alors répété sur une nouvelle image.
On en déduit que s'il existe une certaine régularité dans les images, les filtres vont donc se rapprocher des éléments de base de l'image (ses primitives) et la décomposition sera d'autant plus rapide. Cette règle traduit une évolution vers une représentation éparse de l'image. Cette méthode se rapproche donc de l'algorithme de [Olshausen et Field, 1996], SPARSENET, dans le sens ou en contraignant la représentation de l'image à être éparse, l'algorithme obtenu laisse émerger des filtres du même type. Toutefois l'algorithme d'optimisation (méthode du Gradient Conjugué) est plus analytique, même si les premières version de SPARSENET utilisaient une poursuite gourmande (Olshausen, communication personnelle). Aussi, cet algorithme reste toutefois différent car

il décompose en une somme linéaire des imagettes de la taille des filtres, alors que notre algorithme d'apprentissage décompose l'image entière. Une conséquence importante est que notre algorithme a moins tendance à laisser émerger des filtres similaires avec des centres différents car la décomposition n'a pas lieu sur des imagettes de la taille des filtres, comme dans ces méthodes, mais sur l'image complète. On introduit alors une compétition et la coexistence de filtres similaires par une translation n'est pas stable.
Cette expérience a été menée sur des images naturelles de 128×128, blanchies par le noyau de décorrélation précisé précédemment, et 15 filtres 9×9 (on suit donc exactement le protocole de [Olshausen et Field, 1996]). On a de même utilisé les méthodes de contrôle du gain de SPARSENET pour contrôler la compétition entre les neurones. On vérifie en pratique que si deux neurones convergent vers des filtres similaires alors cette solution sera instable et qu'un seul sera préservé, de telle sorte qu'il n'y a pas de doublons. Aussi après environ 500 itérations, les filtres convergent nettement vers des filtres similaires à ceux observés dans les travaux précurseurs de Linsker [1986], dans les résultats de van Hateren et van der Schaaf [1998] ou dans l'algorithme SPARSENET et qui correspondent aux composantes indépendantes de l'image [Bell et Sejnowski, 1997] (voir Fig. 4.2.6). On a donc prouvé que nos règles plus simples et sans contrainte de type écologique peuvent permettre de retrouver des résultats similaires.

On remarque toutefois que l'ordre des filtres est indifférent et qu'ils n'ont pas de relation spatiale entre eux. On peut alors introduire des relations spatiales excitatrices à courte portée et inhibitrices à portée plus large qui vont stimuler comme dans une carte auto-organisatrice de Kohonen, la formation de relations topographiques. Ici on a expérimenté cette modification de l'algorithme en reliant ces filtres de manière circulaire avec une excitation pour les filtres voisins immédiats et une inhibition (plus précisément un bruitage) à trois filtres de distance. En pratique, on a augmenté la sensibilité des neurones voisins (excitation latérale) pour des cartes voisines (association locale). On voit alors émerger des filtres similaires à l'expérience précédente, mais la colinéarité des lignes dans l'image favorise l'apparition de filtres sélectifs à des orientations voisines sur des cartes voisines (voir Fig. 4.2.7). Toutefois,

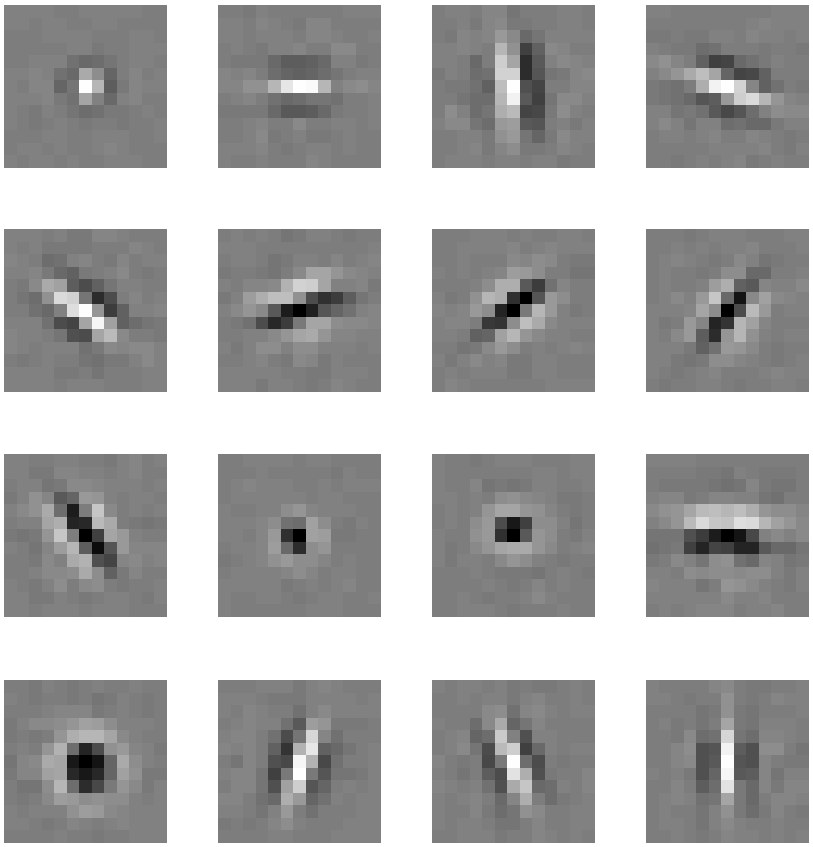

FIG. 4.2.6 – *Émergence de filtres sélectifs à l'orientation.* Dans le cadre de notre code impulsionnel, on a étudié un processus adaptatifs de choix de la base sur-complète. On a choisi un protocole similaire à Olshausen et Field [1996] mais avec une décomposition sur l'image complète et pas seulement sur des imagettes de la taille des filtres. On observe après quelques dizaines d'itération l'émergence de filtres sélectifs à l'orientation comme ce qui est observé dans V1. Comme les images, des scènes d'extérieur, exhibent de nombreuses textures, on voit aussi apparaître des filtres avec des réponses plus complexes (voir filtres 3 et 16, par exemple).

ceux-ci apparaissent plus vite grâce au mécanisme de coopération et surtout forment une certaine continuité qui évoque la formation d'une colonne corticale dans l'aire visuelle primaire. Cet algorithme est à rapprocher des solutions au problème de *séparation aveugle de sources*. En effet, un codage épars peut se rapprocher de l'analyse en composante indépendantes [Comon, 1994] car il tend à rechercher les causes qui ont généré l'image en utilisant de mêmes arguments informationnels de maximisation de vraisemblance (*maximum likelihood* [Olshausen et al., 1998]). Nos résultats sont équivalents à cette méthode, suggérant ainsi que nous avons trouvé par notre algorithme et pour cette taille de dictionnaire les "composantes indépendantes" des images naturelles, *i.e.* les contours orientés des objets.

4.2.3 Conséquences en traitement de l'image

Ce type d'algorithme est inhabituel en traitement de l'image (la Poursuite de Concordance a surtout été appliquée à des signaux temporels comme les sons) et d'autant plus dans le cadre d'un code impulsionnel de la Vision. Il introduit donc de nouvelles techniques d'analyse qui permettent d'envisager de nouvelles applications.

Compression d'image

On connaît l'importance de la compression d'image avec l'explosion de l'utilisation d'images numériques pour des ressources de communication partagées et variables. La qualité des différents standards de compression peuvent être facilement quantifiés par le compromis entre le taux de compression et la qualité de l'image (cela reste donc un critère subjectif) et ils ont connu au cours de leur histoire des périodes de stagnation relatives entrecoupées d'améliorations soudaines liées à des changements de paradigme. Ainsi, les premiers standards comprimaient les données d'intensité des pixels sans se préoccuper de leur position. Une nouvelle ère est symbolisée par le standard JPEG qui code par une technique proche des ondelettes l'image par blocs de 8×8 pixels. Une nouvelle ère semble s'annoncer avec le nouveau standard JPEG2000, basé entièrement sur des techniques d'ondelettes multi-échelles

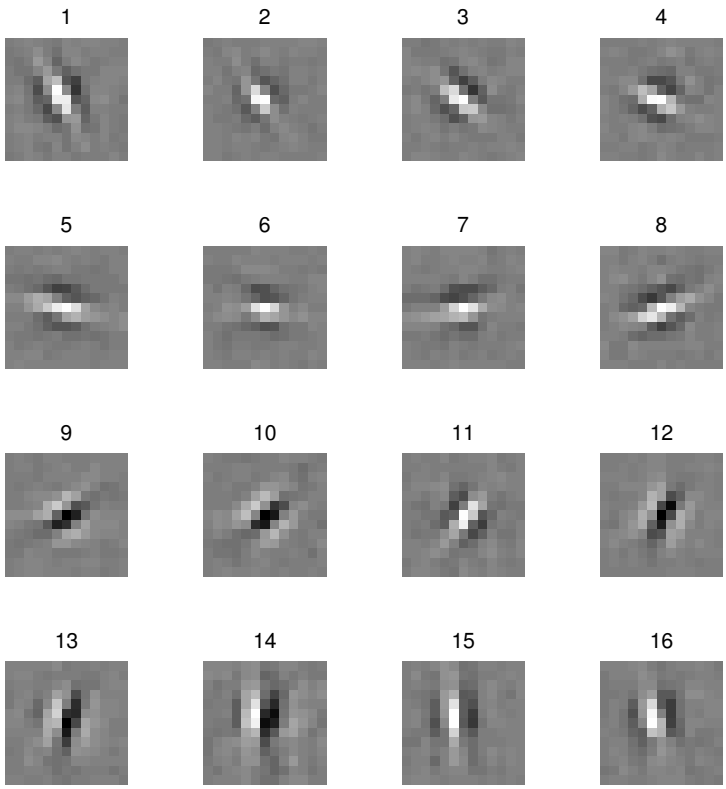

FIG. 4.2.7 – *Émergence d'une carte topologique de filtres sélectifs à l'orientation.* On a étudié l'émergence de filtres corticaux, comme dans Fig. 4.2.6, mais en introduisant des relations spatiales entre les filtres (de numéros voisins circulairement) à la manière d'une carte auto-organisatrice de Kohonen. Pour cela, on a modifié pendant la propagation le seuil des activités voisines pour des filtres voisins (par symétrie des valeurs de contraste, les polarités sont indifférentes). On voit apparaître de la même façon des filtres sélectifs à l'orientation, mais de plus les filtres voisins vont avoir des sélectivités voisines de telle sorte que l'on voit apparaître des sensibilités qui couvrent progressivement toutes les orientations, comme ce qui est observé dans V1 sous forme de *pinwheels*.

et avec le champ encore plus large des représentations sur-complètes.
Afin de comparer la compacité de cette représentation de notre modèle avec d'autres techniques de compression, nous avons codé des images 128×128 pixels par le code visuel de V1 présenté auparavant, et donc sur-complet de 60 fois par pixel. On en déduit dans un premier temps que le nombre de neurones est donc de $128^2.60$ et donc que l'information nécessaire pour coder une impulsion est de $\log_2(128^2.60)/8 \sim 2.48$ byte par impulsion. On peut alors comparer notre modèle au standard JPEG sur ces mêmes images pour les mesures de qualité que nous avons définies. Ainsi, on montre que malgré la simplicité de notre modèle, il rivalise avec ce standard industriel. Cela est d'autant plus vrai pour l'EQM sur les forts taux de compression (JPEG étant optimisé pour une qualité de 75 dans une échelle de 0 à 100) et aussi pour l'EQMB (voir Fig. 4.2.8) à des taux inférieurs de compression. En effet, comme JPEG introduit une analyse par blocs, ceux-ci apparaissent aux forts taux de compression, introduisant une erreur sensible à haute fréquence, auquel l'œil humain — et l'EQMB— sont plus sensibles.

Détection de contours

Un autre application importante en traitement de l'image est la détection des contours des objets. Elle est donc complémentaire dans ce sens à la segmentation des textures des objets et correspond selon le paradigme de Marr à l'esquisse primitive de l'image comme première représentation dans les chemins de la vision. Historiquement, les techniques introduites par Marr et Hildreth [1980] ont été formalisées par Canny [1986] et enfin par Deriche [1987]. Celles-ci ont été enfin à la base d'une théorie de représentation des images par des contours multi-échelles qui est à l'origine de techniques comme la poursuite de concordance [Mallat et Zhong, 1992a].
Les techniques les plus simples calculent le gradient local dans l'image et donnent pour chaque pixel une réponse si l'algorithme considère que ce gradient est maximum localement (ou que la dérivée seconde s'annule localement dans l'algorithme du passage par zéro). Ces algorithmes modélisent donc implicitement un contour par un saut d'intensité, et ces algorithmes introduiront

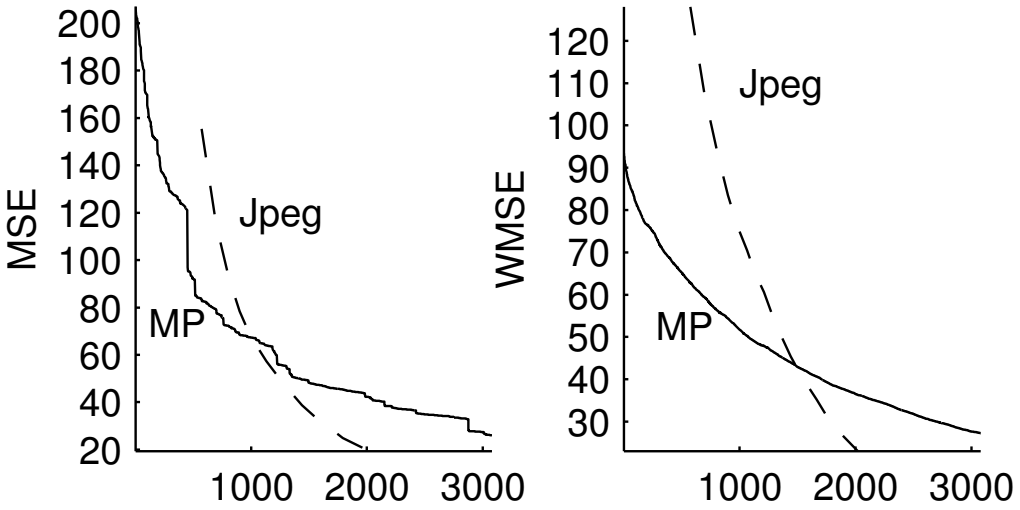

FIG. 4.2.8 – *Compression par le code impulsionnel épars de l'image* Lena On a comparé la qualité de reconstruction en fonction de la taille du code (en $bytes$) Poursuite de concordance (MP) et le standard de compression à différentes qualités (Jpeg). (**Gauche**) EQM (**Droite**) EQMB. Le code visuel est plus compact aux forts taux de compression

---

alors un compromis entre la précision et la localisation du contour. Deriche [1987] a alors dérivé des classes de fonctions qui optimisent ce compromis. Nous allons pour notre part simplement utiliser de tels filtres à différentes orientations et échelles. Dans le cadre d'un code neural, la détection de contours émergera alors de la représentation de l'information par des impulsions. On pourra donc considérer soit les contours comme l'ensemble des points ayant tiré ou alors garder la représentation complète comme la forme ou le type de contour qui sera propagée en avant.

Expérimentalement, on a appliqué une telle technique à une image de synthèse choisie pour la difficulté liée à la détection de ses contours. On a comparé les résultats de la propagation progressive des impulsions par rapport à l'algorithme de Canny pour différents seuils qui laissent apparaître un même taux de points activés dans l'image. Notre méthode possède le premier avantage de pouvoir définir des classes différentes de contours et donc de transmettre par cette "esquisse" plus d'information que l'image des contours par l'algorithme

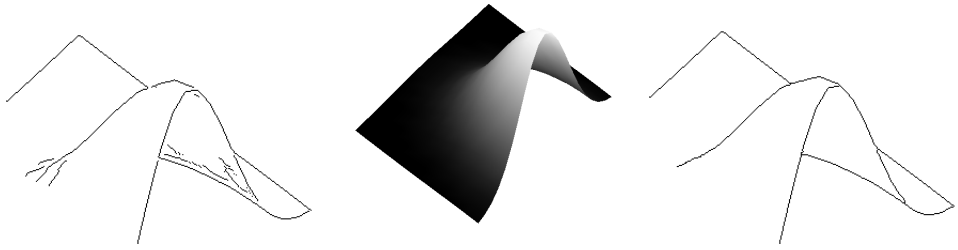

FIG. 4.2.9 – *Détection de contours.* **(Milieu)** Une image de synthèse présente une surface plissée dont nous cherchons les contours. La détection est aisée sur la surface, mais **(Gauche)** la méthode de Canny présente certains problèmes aux faibles contrastes mais surtout aux coins aigus et aux jonctions (algorithme implanté par MATLAB©). **(Droite)** La stratégie de connaissance a priori garanti la propagation en avant des contours colinéaires ou cocirculaires. La représentation obtenue par le codage impulsionnel permet de découpler les contours suivant leur paramètres (orientation, taille) et permet d'éviter les erreurs lors de croisements de lignes de contour (on a représenté les pixels qui correspondent à une impulsion ainsi que les lignes qui correspondent à des liens entre contours voisins).

---

de Canny ; cette méthode peut alors servir comme base à la reconstruction de l'image [Mallat et Zhong, 1992a,b]. De plus on observe que les contours sont détectés suivant des avalanches se propageant de proche en proche le long des "crêtes" de l'activité. En définissant cet algorithme par un jeu de règles locales qui imitent les mécanismes de perception du contour [Peterhans et Von der Heyt, 1989], on définit ce processus dans son implantation corticale de façon analogue aux automates cellulaires.

4.3 Systèmes visuels dynamiques et applications

> Above all, the cerebral cortex is a part of a multi-level biological system which has a global coherence.
>
> Burnod [1989, p.7]

Le code visuel impulsionnel épars défini pour l'aire visuelle primaire peut être étendu à des systèmes visuels complets. Nous allons ici essayer de définir un modèle général de traitement cortical, pour pouvoir l'étendre à un modèle multicouches et l'appliquer à des expériences réelles.

4.3.1 Heuristiques attentives et génération de cartes d'attention

Tout d'abord, nous avons évoqué précédemment la possibilité de modifier le critère de choix afin de modifier la propagation des impulsions et donc la représentation de l'image. Nous allons voir maintenant les conséquences du choix d'heuristiques simples sur ce critère et son lien avec la génération de cartes attentives.

Heuristiques de propagation

En effet, en plus des méthodes que nous avons introduites pour choisir à un moment donné le filtre le plus adapté, on peut introduire des *heuristiques* sous forme de règles utilisant les filtres déjà choisis. On peut à ce titre comparer ces règles aux saccades des yeux et qui vont être induites par les points saillants de l'image. De façon pratique, on peut dans un premier temps définir une distance $\|i-j\|_{\mathcal{C}}$ qui reflète la dissemblance entre des filtres i et j de la carte $\mathcal{C}$ et introduire cette distance au critère de choix

$$i^t \quad = \quad \mathrm{ArgMax}_{i\in\mathcal{D}}(|C_i^t| + \lambda\|i - i^{t-1}\|_{\mathcal{C}}) \tag{4.3.1}$$

où λ peut être considéré comme un terme régularisant qui va contrôler l'influence de cette heuristique. On introduit donc un compromis dans le choix d'un filtre entre la saillance d'un objet et la "distance" des choix successifs. Cette règle et alors très facile à incorporer au système de code épars que nous avons implanté. Dans le cadre du traitement automatique d'images de grandes tailles (comme les images satellitaires ou de mammographie) on a souvent besoin de détecter des structures dans des zones bruitées. Dû au caractère parallèle du traitement dans le code impulsionnel, une solution est

de concentrer la propagation sur une zone de l'image "saillante", déterminée par un processus plus simple.
On a expérimentalement utilisé une distance s'exprimant seulement en fonction de la distance entre les centres des filtres, et donc ici sur la distance depuis le dernier filtre choisi. Cette méthode a été appliquée à des images de visages d'une base donnée [Samaria et Harter, 1994]. On a ensuite synthétisé de nouvelles images en superposant les visages détourés manuellement sur un bruit spatial uniforme. Si on a une connaissance a priori de la position du visage, on peut influencer le choix des filtres en baissant le seuil des neurones centrés sur cette zone (voir Fig. 4.3.1-(Milieu)), une stratégie utilisée dans [Van Rullen et Thorpe, 1999, 2002]. Par contre, quand cette position n'est pas connue a priori, l'image du visage va dans ce cas induire les premiers tirs qui donneront une information non-supervisée des parties saillantes de l'image et favoriser le tir des filtres voisins. On voit alors les parties correspondant à des zones cohérentes se propager en premier, et l'avalanche de tirs locaux entraîne la propagation "en avant" du visage.

Carte attentive dynamique de saillance

Cette expérience se rapproche donc dans un sens de la carte attentive de saillance de Itti et al. [1998]. Celle-ci consiste en un empilement de cartes neurales modélisant des neurones de LIP dont on suppose qu'ils collectent les données saillantes de l'image (comme le contrastes en luminosité, en chrominance, direction et leurs combinaisons) pour diriger le mouvement des yeux. Ces données saillantes se distinguent d'autres caractéristiques de l'image dans le sens où elles "sautent aux yeux" (*pop-out*) au lieu de nécessiter une exploration sérielle dans l'image (comme par exemple la recherche la recherche d'un garçon avec un maillot rayé rouge dans une foule). La méthode de Itti et al. [1998] collecte ces différents aspects dans l'image pour les intégrer et créer une carte attentive qui correspond à une mesure quantitative de la saillance. L'algorithme choisit alors récursivement le maximum vers lequel le regard est porté et inhibe en retour (*inhibition of return*) une aire de surface fixe la surface de saillance (voir Fig. 4.3.2). Cet algorithme jouit d'une grande

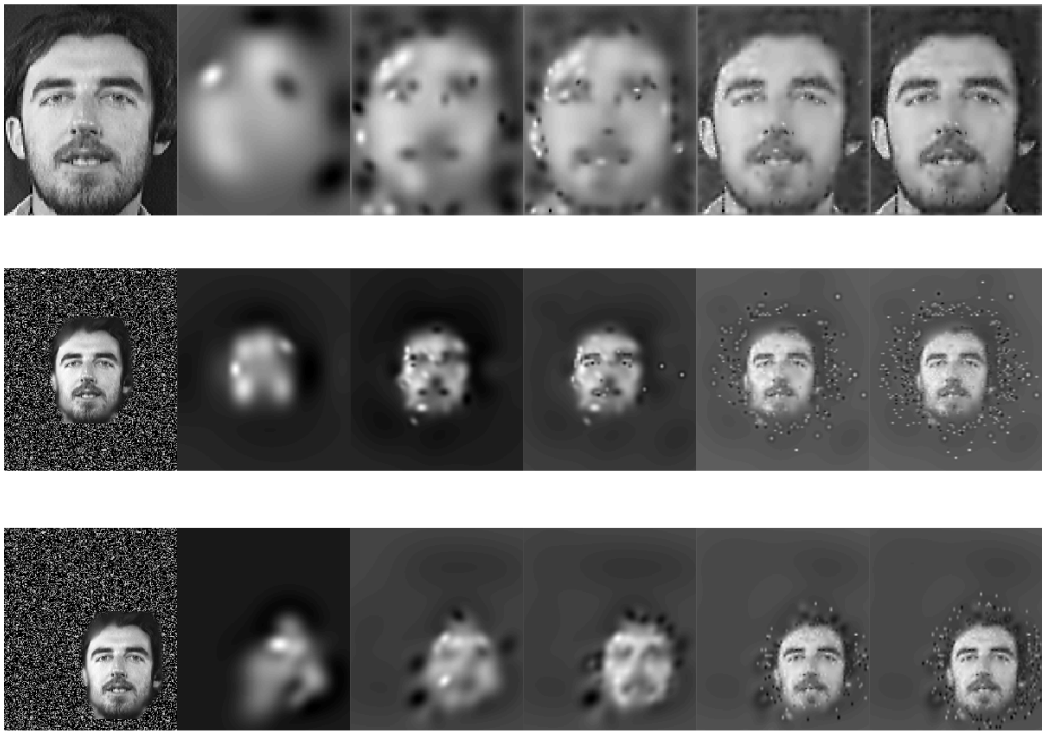

FIG. 4.3.1 – *Heuristiques de propagation.* On a appliqué la poursuite de concordance à (**Gauche**) l'image 1 du sujet 11 de la base de donnée d'image ATT [Samaria et Harter, 1994], et (**Haut**) la reconstruction est présentée pour resp. de gauche à droite 10, 50, 100, 450 et 750 impulsions. (**Milieu**) Si on applique une heuristique favorisant les neurones centraux, l'information du centre de l'image est propagée en avant (Modulation attentive *top-down*). (**Bas**) Dans le cas où l'heuristique favorise des tirs de filtres voisins, alors le visage —superposé à un bruit de fond Gaussien— et qui provoque par ses structures les premiers tirs provoque une "avalanche" de tirs correspondant au visage, de sorte que celui-ci est propagé en avant (Modulation attentive *bottom-up*).

popularité et est promis à de nombreuses applications.

On comprend alors le lien avec la poursuite de concordance, et l'apport que celle-ci peut induire. En effet, l'algorithme de carte de saillance souffre du caractère arbitraire de l'inhibition de retour et un point de saillance peut se retrouver dans "l'ombre" d'un point précédent. Nous avons donc imaginé un système avec des filtres représentant des points saillants (des orientations, des coins) et utilisé l'algorithme précédent en utilisant une distance qui favorise des configurations saillantes, comme des orientations orthogonales. Ces points peuvent dans le système visuel se propager via les couches corticales V et VI respectivement commander la position (par les saccades oculaires) et amplifier la carte thalamique.

On n'obtient donc pas une carte attentive statique comme pour la méthode d'Itti, mais une carte dynamique qui va progressivement soustraire de l'image les caractéristiques locales : il va donc produire une suite de points de saillance qui va entraîner une coopération locale, comme pour l'heuristique locale. Toutefois, l'inhibition par contre dans ce cas sera exacte et on pourra observer la zone de l'image correspondante qui est propagée dans ce phénomène attentif. Nous avons réalisé cette algorithme sur une image choisie pour présenter des objets (ici des bateaux, voir Fig. 4.3.2) sur plusieurs plans. On a dans un premier temps tenu seulement compte des contours orientés car nous nous limitons toujours à des images en niveaux de gris. On voit sur la carte de saillance équivalente –qui mesure l'énergie cumulée sur toutes les couches— que l'énergie est effectivement concentrée sur ces objets, et qu'au cours de la propagation ces zones sont progressivement atténuées par les interactions latérales. L'image correspondante est progressivement reconstruite en passant d'un bateau en premier plan aux bateaux plus lointains et puis au fond de l'image.

4.3.2 Code impulsionnel multicouche : vers des représentations corticales

Malgré la grande souplesse de ce codage et les applications dont nous venons de faire une revue, ces algorithmes supposent implicitement la re-

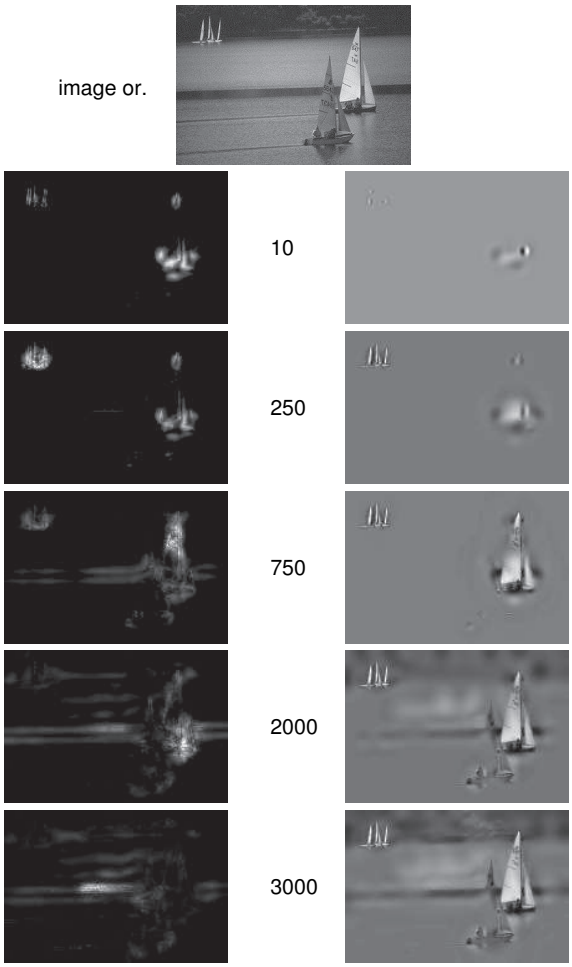

FIG. 4.3.2 – *Carte de saillance dynamique.* L'algorithme de carte de saillance permet de générer une suite de points de saillance décroissante en se basant sur des critères locaux. Cet algorithme a été ici appliqué à (**Haut**) une image de la base d'Itti et al. [1998]. On peut déduire de la répartition de l'énergie correspondant aux coefficients (**Gauche**) une carte de saillance dynamique indiquant (en clair) les points correspondant à la plus haute énergie. (**Droite**) L'image propagé par ce processus montre la reconstruction de l'image avec ce processus. Cette expérience exhibe des résultats similaires, mais permettent de tenir compte de points de saillances à plusieurs échelles. Aussi, l'inhibition de retour est exacte et on peut reconstruire l'image correspondant à cette propagation.

construction analogique intermédiaire d'une image, qui génère alors une propagation dans la nouvelle couche neurale. Nous allons alors étendre notre algorithme à un nombre arbitraire de couches et qui ne nécessite pas une reconstruction intermédiaire de l'image.

Formalisation

Plaçons-nous dans un premier temps dans une architecture où les impulsions de la rétine arrivent directement sur l'aire visuelle primaire selon une suite d'indices $\{i^t\}$ correspondant à des filtres rétiniens ϕ_i. On peut alors d'abord imaginer que l'image est reconstruite progressivement par $L(t+1) = L(t) + p^t.\frac{m^t}{{N_{i^t}}^2}.\phi_{i^t}$ où p^t est la polarité et m^t est la modulation (LUT) calculée comme précédemment. Si l'on note ψ_j les filtres de V1, on peut itérativement calculer l'activité d'un neurone j de V1 par

$$C'_j(t+1) =< L(t+1), \psi_j >= C'_j(t) + p^t.m^t < \frac{\phi_{i^t}}{{N_{i^t}}^2}, \psi_j > \tag{4.3.2}$$

Cette approche se rapproche du modèle proposé par Hubel et Wiesel où les filtres de V1 sont exprimés comme une composition de filtres rétiniens. On en déduit donc que l'activité est calculée directement depuis l'adresse des impulsions, sans passer par une reconstruction.
En définissant un seuil θ, on peut alors définir un nouveau mécanisme impulsionnel dans V1. Comme le filtre qui tire sera le filtre qui correspondra temporairement à la meilleure concordance, on pourra construire des inhibitions de la même manière et mettre à jour les activités voisines au tir du neurone $j^{t'}$, en retranchant la projection de $\psi_{j^{t'}}$:

$$C'_j(t') \leftarrow C'_j(t') - C'_{j^{t'}}(t').\frac{< \psi_{j^{t'}}, \psi_j >}{\|\psi_{j^{t'}}\|^2} \tag{4.3.3}$$

On peut alors définir la propagation dans de nouvelles couches par ce formalisme et imaginer aussi que ces propagations suivent des architectures copiant la complexité du réseau neural cortical, mais nous nous limiterons ici à des architectures hiérarchiques (en avant) pour conserver toutes les propriétés de la poursuite de concordance. On en déduit que dans cette formalisation pure-

ment événementielle d'automates déterministiques, le temps est basé sur les événements et en premier lieu sur les impulsions rétiniennes. Par conséquent, si d'un coté l'ensemble des temps de propagation d'une couche sont inclus dans l'ensemble des temps de tir de la couche suivante, on peut aussi voir que chaque tir dans une couche supérieure est due à une avalanche d'impulsions depuis une impulsion rétinienne.

Vers un modèle fonctionnel de colonne corticale

Afin de synthétiser les résultats que nous avons présenté et proposer un modèle fonctionnel, nous allons maintenant décrire un modèle de colonne corticale comme un automate adaptatif [Burnod, 1989]. Ce modèle théorique va donc comprendre plusieurs neurones liés fonctionnellement que l'on va décrire au niveau de description de l'aire corticale comme un système autonome. Cette colonne va être caractérisée par son entrée et sa sortie impulsionnelles et par son état interne. Par analogie avec les observations neurophysiologiques et comme décrit dans Sec. 2.1.3, cet automate sera constitué de plusieurs couches que nous simplifierons en ne considérant que les entrées et sorties thalamiques et cortico-corticales.
Une couche d'entrée intègre les entrées thalamiques et cortico-corticales — c'est-à-dire dans le cadre de notre modèle simplifié de V1 les impulsions générées par le code rétinien— pour définir l'activité analogique de la colonne. Cette activité est transformée en un événement d'autant plus vite que cette activité est forte : c'est la voie *en avant* de la colonne. Cet événement est alors transmis par des connexions latérales et symétriques (ce qui correspondrait dans un modèle biologique à la couche II) pour rendre compte de l'événement gagnant sur les voisins de telle sorte à ce que la représentation soit éparse et efficace : c'est la voie *latérale* de la colonne. D'après notre formalisation, les poids correspondant peuvent être déterminés par la corrélation des filtres et donc par une règle de type hebbienne.
On définit alors une couche superposée mais dans la direction descendante, qui va correspondre à la voie *en arrière* de la colonne et qui va implanter fonctionnellement les processus régulateurs et attentifs. On définira dans ce

cadre une activité correspondant à l'*attente* de la colonne intégrée depuis les événements des voies en avant des colonnes voisines. Cette activité d'attente de la colonne peut se propager sous forme d'un événement en retour vers les sources thalamiques et servir de régulateur du flux d'information visuelle. Elle influence aussi directement la sensibilité de la voie en avant en modifiant le seuil de génération d'un événement en avant depuis l'activité de la colonne. On pourra dans ce cadre définir les poids liés à cette activité par un processus de renforcement qui, par exemple dans V1, mettra en avant les contrastes d'orientation.

Modélisation de V4

Afin d'illustrer ce modèle de façon fonctionnelle, imaginons que nous connections une première aire visuelle sélective aux orientations locales (soit comme V1) à une aire composée de neurones sélectifs à des angles, courbures et lignes (et qui correspondrait donc biologiquement à V4). Pour des raisons de volume de calcul, nous ne considérerons ici plus qu'une fraction de la carte rétinotopique. Les événements calculés depuis V1 se projetteront sur ces neurones de telle sorte qu'à chaque point correspondra une activité d'autant plus forte que la *courbure locale*, c'est-à-dire que l'angle entre des segments contigus est proche d'une pointe.
Si l'on représente une reconstruction de l'information événementielle depuis les sorties de ces deux aires, nous obtiendront des résultats dynamiquement différents. Comme nous l'avons vu, la sortie de V1 met en avant les contrastes orientés dans l'ordre de la mesure de leur singularité, et nous verrons typiquement apparaître dans une image les bords des objets, formant ainsi une esquisse primitive de l'image. La sortie de notre modèle de V4 représentera donc progressivement les angles aigus (flèches, coins, courbes) puis les lignes et les détails. En particulier, cette aire va pouvoir détecter des jonctions de type "T", c'est-à-dire des croisements de contours caractéristiques d'occlusions.
Aussi, les événements correspondant à l'attente et l'activation des colonnes de V4 peuvent être rétropropagées sur V1 puis sur le CGL pour modifier le

flux d'information. Si l'on prend cette propagation en compte, on observe alors que dès V1, l'information saillante correspondant aux coins peut être propagée en avant grâce à l'information de V4. Dans un cadre biologique plus large, on peut imaginer que les vagues d'activités plus lentes que nous avons évoqué (voies Parvo et Konio, voir Sec. 2.2.2) vont alors confronter cette information précise avec l'information descendante plus imprécise de la voie Magno qui aura contrasté par les processus attentifs les voies saillantes, selon un modèle *itératif* [Bullier et Nowak, 1995].

4.3.3 Applications à la robotique et à la vision active

Comme nous l'avions évoqué au début de la thèse, la vision se définit par les buts et les contraintes qui lui sont imposées. Nous avons donc essayé d'implanter des paradigmes issus des études émergeant de cette nouvelle approche du code de la vision à des systèmes robotiques.

Couplage multisensoriel

Nous avons tout d'abord étudié un système perceptif multi-sensoriel initié en France au LORIA par D. Martinez [Rochel et al., 2002]. Ce système consiste en un robot mobile —de type Koala— dont le but est de pouvoir naviguer dans un environnement partiellement connu en utilisant des indices visuels et olfactifs et est spécialement conçu pour pouvoir être utilisé en milieu industriel lors d'incidents chimiques par exemple, où une action non-humaine est nécessaire. Des capteurs olfactifs de type Figaro sont placés des deux cotés du robot en face de tubes aspirant l'air par des pompes, de telle façon qu'on obtient un signal analogique multi-canal et stéréo de l'environnement chimique. Toutefois, ces capteurs manquent de sensibilité temporelle et olfactive, particulièrement lorsque le robot est en mouvement et un traitement par des techniques neurales est alors utilisé pour catégoriser les odeurs.

Les capteurs olfactifs sont particulièrement peu performants si la source de l'odeur est éloignée du robot, d'autant plus si cette odeur —comme l'éthanol— est volatile. Nous avons donc considéré la possibilité d'utiliser d'explorer l'environnement, dans ce cas où aucune odeur n'est catégorisée, en

FIG. 4.3.3 – *Système olfacto-visuel Nose.* Ce robot, de type *Koala*, est équipé d'une tourelle de vision stéréoscopique mais aussi de capteurs stéréoscopiques d'odorat, "reniflant" la nature du gaz aspiré par les tubes latéraux. Il a été utilisé pour réaliser une expérience de couplage de la vision et de l'odorat. La vision sert alors de système de guidage non-supervisé grâce à un processus attentif alors que l'odorat permet de catégoriser des odeurs présentes localement. Ce système est construit pour pouvoir être appliqué dans des interventions automatisées lors d'incidents de type chimique.

utilisant la perception visuelle. Grâce à l'aide de Dirk Walther, du laboratoire de Christof Koch à Caltech, Pasadena, lors de l'École d'Été d'Ingénierie Neuromorphique de Telluride, nous avons utilisé le modèle de carte de saillance de [Itti et al., 1998], que nous avons décrit plus haut. Ceci a permis d'extraire de l'image prise par le robot une zone de plus forte saillance qui donnait au robot une direction vers laquelle il avançait. Le robot s'arrêtait après une courte distance pour aspirer l'air environnant jusqu'à ce que la réponse aux senseurs donnée par le réseau de neurones impulsionnel de catégorisation soit assez précise auquel cas l'expérience était considérée comme réussie.
L'expérience commandait en temps réel depuis Telluride, Colorado le robot situé en Lorraine par un protocole TCP/IP. Le robot était placé dans un environnement classique à 4 m environ de la source (de l'éthanol) qui était diffusée par un ventilateur. Utilisant une heuristique similaire à la descente de gradient, les capteurs ne détectent à cette distance aucune odeur et le robot s'approchait donc vers le point le plus saillant. Or le nombre d'objets saillants est limité et le robot s'approche éventuellement vers la source. Toutefois, la diffusion de l'odeur est très turbulente et nous n'avons observé une catégorisation que quand le robot se retrouvait à seulement 50 cm de la source auquel cas le robot pouvait se guider de façon bisensorielle par la vision et les capteurs stéréo-olfactifs.

Détection d'un objet connu

Nous avons enfin appliqué les principes de codage multi-couche impulsionnel à une tâche simple de détection d'un objet connu à des fins de guidage. Le robot est un véhicule mobile développé au laboratoire d'Informatique et d'automatique de Supaéro, Toulouse. Ce robot autonome comporte une unité de calcul et est contrôlé par un calculateur dédié. Sa double caméra produit une image vidéo stéréoscopique de 192×144 pixels sur 8 bits de niveaux de gris. Le robot profite des développements dans le laboratoire sous forme de librairies des différentes fonctions que l'on désire implanter. Notre but est donc là de fournir un algorithme de détection d'un balise connue (voir Fig. 4.3.4) à différentes tailles, orientations, distances et configurations.

Pour réduire les calculs, nous avons implanté un système de trois couches rétinotopiques de résolutions décroissantes correspondant grossièrement à la rétine, à V1/V4 et à IT. La rétine produit une représentation des contours, comme une esquisse de l'image. La couche V1/V4 correspond à une couche de filtres sélectifs à des orientations suivant les directions de contours présentes dans les balises (verticales et diagonales) et aux coins correspondant aux balises. Ces filtres ont été appris de manière non-supervisée sur des images de balises sans distracteur et ont été étendus à l'architecture multi-échelle. On a alors ajusté les angles compatibles avec le triangle pour que ceux-ci soient propagés en premier. À l'image du triangle de Kanisza, la détection est réalisée quand 3 angles compatibles sont propagés.

La dernière couche a été apprise de façon supervisée grâce à une base d'images de balises marquées. Les résultats sont sans surprise car les contours des balises, par leur fort contraste, sont propagées en premier dès la rétine. On voit alors apparaître rapidement les structures linéaires de l'image et la réponse de IT. Comme ce système est utilisé pour une navigation, on peut donc grâce aux capteurs odométriques prévoir la direction, taille et orientation de la balise depuis son estimation précédente. Cette technique a été utilisée grâce à une régulation "du haut en bas" de l'attention qui permet de réduire énormément les coûts de calculs. Ceci permet d'envisager une implantation sur le robot en décomposant la phase lente de recherche sérielle de la cible dans l'environnement puis une phase de guidage attentif. Elle ouvre ainsi la porte à l'implantation d'un guidage d'un robot mobile par des amers qui est le sujet de la thèse d'Emmanuel Zenou dans ce même laboratoire.

FIG. 4.3.4 – *Détection de balises par un robot.* Cette image présente une scène du robot mobile contenant les balises utilisées, des "drapeaux" triangulaires noirs sur fond blanc. Cette simulation se rapproche du traitement d'une image dans le système visuel : l'information est pré-traitée à travers un système d'une couche simulant l'aire V1, puis par une couche similaire à V4 détectant des angles aigus (en particulier ceux correspondants aux polarités des balises). Une couche similaire à IT détecte enfin en combinant cette dernière information les différents types de triangles rencontrés (différentes positions, distances et angles). La propagation de l'information est progressive et hiérarchique dans le système.

Conclusion

> I am never content until I have constructed a mechanical model of the subject I am studying. If I succeed in making one, I understand ; otherwise I do not.
>
> *Lord Kelvin*

Nous éloignant de la présentation de la Vision comme une méthodologie de traitement statique de l'image, nous avons décrit les processus visuels comme un flux dynamique, codé temporellement et de façon asynchrone, qui permet d'interagir avec le monde. Nous avons décrit plus particulièrement le traitement ultra-rapide comme un aspect de la Vision pour lequel l'information visuelle est codée par des vagues successives d'impulsions se propageant dans l'architecture du système visuel et où l'information "importante", *saillante*, est progressivement mise en avant. En particulier, nous avons vu que cette description permet de définir un code fonctionnel de transmission et de traitement de l'information visuelle par l'ordre d'activation de ces impulsions. Aussi, nous avons alors mis en avant l'importance des statistiques des stimuli naturels et des lois écologiques qui permettent de prévoir l'adaptation des substrats biologiques du calcul neuronal.
Au cours de ce travail, nous nous sommes efforcés d'intégrer ce travail dans le cadre des études menées à Toulouse en collaboration par Manuel Samuelides au CERT-ONERA et par l'équipe de Simon Thorpe au CERCO. En particulier, les contributions personnelles à ces travaux consistent en des études sur plusieurs niveaux de description.

- au niveau sub-neuronal, l'étude fonctionnelle du comportement de modèles de neurones réalistes et en particulier l'étude de l'adaptation de la dynamique à la statistique des entrées et l'émergence non-supervisée de sélectivité extrayant la cohérence d'un flux asynchrone d'impulsions.
- au niveau neuronal, le lien entre ces modèles détaillés et un code visuel complet basé sur l'ordre d'activation des impulsions. En particulier, on a étudié des modèles d'adaptation mais aussi une formalisation permettant de prévoir le comportement de tels neurones et des modèles

appliquant ces modèles à des applications visuelles, comme le traitement rétinien.

– au niveau des assemblées de neurones, on a défini une méthode permettant d'implanter —grâce à des interactions latérales— une représentation impulsionnelle sur-complète et éparse. Nous avons alors montré son efficacité, tant au niveau de la capacité de représentation que des applications pour le traitement, mais aussi sa généralité en l'étendant à plusieurs couches et en définissant des processus adaptatifs. Enfin, on a relié des heuristiques de propagation à des phénomènes attentifs qui peuvent modéliser certains aspects de l'attention visuelle.

Ces contributions s'appuient donc sur les paradigmes d'un mouvement émergent du connexionnisme en Sciences Cognitives qui établit les bases d'une théorie biophysique des calculs cognitifs en mettant en avant les aspects asynchrones du calcul neural. Les nouvelles méthodes de traitement du signal et de l'image que nous avons introduites se basent donc sur ces principes biologiques et se veulent comme des ouvertures vers l'implantation de systèmes visuels efficaces. En particulier, l'implantation matérielle de ces algorithmes permettra le passage de la simulation sur un calculateur synchrone à des expérimentations qui soulèveront des problèmes qui sans nul doute nous éclaireront sur le fonctionnement de la Vision. En particulier, la définition actuelle d'un standard de protocole *Address-Event Representation* doit se faire en prenant en compte l'aspect dynamique des calculs et l'importance grandissante des calculateurs asynchrones Sutherland et Lexau [2001].

En parallèle, ces théories permettent de fonder progressivement une approche scientifique des aspects les plus subjectifs et intimes du cerveau. Si notre contribution se limite à des apprentissages simples, des règles plus avancées sont nécessaires pour comprendre d'autres processus cognitifs, comme le langage naturel. Les progrès de l'imagerie nous aident alors à comprendre ces mécanismes, mais nous ne sommes qu'à l'aube d'une série de bouleversements de notre vision de la pensée.

Abréviations et acronymes

Nous présentons ici les abréviations et acronymes utilisés en pointant sur la section où ceux-ci sont initialement définis.

SNC	Système Nerveux Central	Sec. 1.1.1
PA	Potentiel d'Action	Sec. 2.1.2
PPS	Potentiel Post-Synaptique	Sec. 2.1.2
EEG	Électro-encéphalogramme	Sec. 2.1.3
CG	Cellules Ganglionnaires	Sec. 2.2.1
PhR	Photorécepteurs	Sec. 2.2.1
CGL	Corps Genouilé Latéral	Sec. 2.2.2
V1	Aire visuelle primaire	Sec. 2.2.2
IT	Inféro-Temporal	Sec. 2.2.2
CUR	Catégorisation Ultra-Rapide	Sec. 2.3.1
LIF-SC	Leaky-Integrate & Fire - Synaptic Current	Sec. 3.1.1
STDP	Spike-Time Dependant Plasticity	Sec. 3.1.2
LUT	Look-Up-Table	Sec. 3.2.2
EQM	Erreur Quadratique Moyenne	Éq. 3.2.10
IM	Information Mutuelle	Éq. 3.2.10
EQMB	Erreur Quadratique Moyenne Balancée	Éq. 3.2.22
MP	Matching Pursuit	Sec. 4.1.2

Publications

L. Perrinet et M. Samuelides *A generative model for spike time dependent hebbian plasticity.* In *Proceedings of DYNN, Bielefeld.*, 2000.

L. Perrinet, A. Delorme, S. J. Thorpe et M. Samuelides *Network of integrate-and-fire neurons using rank order coding A : how to implement spike timing dependant plasticity.* Neurocomputing p. 817–22, 2001.

A. Delorme, L. Perrinet, S. J. Thorpe et M. Samuelides *Network of integrate-and-fire neurons using rank order coding B : spike timing dependant plasticity and emergence of orientation selectivity.* Neurocomputing p. 539–45, 2001.

L. Perrinet et M. Samuelides *Coherence detection in a spiking neuron via hebbian learning.* Neurocomputing, p. 133–9, 2002*a*.

L. Perrinet et M. Samuelides *Sparse image coding using an asynchronous spiking neural network.* In *Proceedings of ESANN*, pages 313–18, 2002*b*.

L. Perrinet et M. Samuelides *Visual strategies for sparse spike coding.* In *Actes de Neurosciences et Sciences de l'Ingénieur, L'Agelonde*, 2002*c*.

L. Perrinet, M. Samuelides et S. Thorpe *Sparse spike coding in an asynchronous feed-forward multi-layer neural network using Matching Pursuit.* Neurocomputing, in press, 2002.

Bibliographie

Larry F. Abbott et Sacha B. Nelson. Synaptic plasticity : taming the beast. *Nature Neuroscience*, 3 :1178–83, 2000.

Larry F. Abbott, Juan A. Varela, Kamal Sen, et Sacha B. Nelson. Synaptic depression and cortical gain control. *Science*, 275 :220–4, 1997.

Edgar Adrian. *The Basis of Sensation : The Action of Sense Organs.* London : ChristoPhers., 1928.

D. Alleysson, S. Susstrunk, et J. Hérault. Color demosaicing by estimating luminance and opponent chromatic signals in the fourier domain. In *Color Imaging Conference, CIC10*, 2002.

Luis Alvarez, Yann Gousseau, et JeanMichel Morel. The size of objects in natural images. *CMLA*, 1999.

Shun-ichi Amari. Theory of adaptive pattern classifiers. *IEEE Trans. Electr. Comput.*, 16 :200–307, 1998.

M.A. Arbib. *The handbook of Brain Theory and Neural Networks.* The MIT Press, Cambridge, MA, 1995.

Joseph J. Atick. Could information theory provide an ecological theory of sensory processing ? *Neural Computation*, 3(2) :213–52, 1992.

Joseph J. Atick et A. Norman Redlich. What does the retina know about natural scenes ? *Neural Computation*, 4(2) :196–210, 1992.

P. Bach-y Rita. Tactile vision substitution : past and future. *Int. J. Neurosci.*, 1-4(19) :29–36, 1983.

Roland Baddeley. The correlational structure of natural images and the calibration of spatial representations. *Cognitive Science*, 21(3) :351–372, 1997. URL `http://citeseer.nj.nec.com/baddeley96correlational.html`.

Wyeth Bair et Christof Koch. Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Computation*, 8 (6) :1185–1202, 1996.

Philippe Barbe et Marc Hallin. Statistiques de rang linéaires ; normalité asymptotique et théorème de projection de Hàjek. In *Inférences non paramétriques.* Association pour la Statistique et ses Utilisations, 1995.

Horace B. Barlow et P. Foldiak. Adaptation and decorrelation in the cortex. In G Mitchinson. ed. R Durbin, C Miall, editor, *The Computing Neuron*, number 4, page 54–72. New York : Addison-Wellesley, 1989.

Anthony J. Bell et Terrence J. Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37(23) :3327–38, 1997.

R. Ben-Yishai, R. L. Bar-Or, et H. Sompolinsky. Theory of orientation tuning in visual cortex. *Proc. Natl Acad. Sci. USA*, 92 :3844–8, 1995.

G-Q Bi et M-M Poo. Synaptic modifications in cultured hippocampal neurons : Dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 18 :10464–72., 1998.

I. Biederman. Human image understanding : Recent research and a theory. *Computer Graphics, Vision and Image Processing*, 32 :29–73, 1985.

Kwabena A. Boahen. The retinomorphic approach : pixel-parallel adaptive amplification, filtering, and quantization. In *Analog Integrated Circuits & Signal Processing*, volume 13, pages 53–68. 1997. URL `http://www.neuroengineering.upenn.edu/boahen/boahen_body.htm`.

Kwabena A. Boahen. Point-to-point connectivity between neuromorphic chips using address-events. *IEEE Transactions on Circuits and Systems, Invited Tutorial.*, 2000. URL `http://www.neuroengineering.upenn.edu/boahen/boahen_body.htm`.

L.J. Borg-Graham, C. Monier, et Y. Fregnac. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, 6683 (393) :369–73, 1998.

Lyle J. BorgGraham. Interpretations of data and mechanisms for hippocampal pyramidal cell models. In *Cerebral Cortex*, volume 13. P. S. Ulinski, E. G. Jones and A. Peters, New York : Plenum Press, March 1999.

Nicolas Brunel et Simone Sergi. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, 1999.

Robert W. Buccigrossi et Eero P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. In *Proceedings ICASSP-97 (IEEE International Conference on Acoustics, Speech and Signal Processing)*, number 414, Munich, Germany, 1997. URL `http://citeseer.nj.nec.com/article/buccigrossi97image.html`.

J. Bullier et L. G. Nowak. Parallel versus serial processing : new vistas on the distributed organization of the visual system. *Curr Opin Neurobiol*, 4 (5) :497–503, 1995.

A. N. Burkitt et G. M. Clark. New technique for analyzing integrate and fire neurons. In *Neurocomputing*, volume 26-27, pages 93–99, 1999.

Yves Burnod. *An Adaptive Neural Network : The Cerebral Cortex.* Masson, 1989.

Alvin Burt. *Textbook of neuroanatomy.* Saunders, 1992.

Peter J. Burt et Edward H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, COM-31,4 :532–40, 1983. URL `http://citeseer.nj.nec.com/burt83laplacian.html`.

Santiago Ramòn Y Cajal. *Histologie Du Système Nerveux de l'Homme et Des Vertébrés.* Maloine, Paris, 1911.

J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 8 :679–98, 1986.

Eric Castet et Guillaume S. Masson. Motion perception during saccadic eye movements. *Nature Neuroscience*, 3(2) :177–83, 2000.

Hideyuki Cateau et Tomoki Fukai. A stochastic method to predict the consequence of arbitrary forms of spike-timing-dependent plasticity. *Neural Computation*, 2002.

S. Celebrini, Simon J. Thorpe, Y. Trotter, et M. Imbert. Dynamics of orientation coding in area V1 of the awake primate. *Vis Neurosci*, 5(10) :811–25, 1993.

Central Bureau of the Commission Internationale de L'Éclairage (Vienna, Austria). Colorimetry, second edition. *Signal Processing*, 15.2, 1986.

F.S. Chance, Sacha B. Nelson, et Larry F. Abbott. Synaptic depression and the temporal response characeristics of v1 simple cells. *The Journal of Neuroscience*, 18 :4785–99, 1998.

Gal Chechik et Naftaly Tishby. Temporal dependent plasticity : An information theoretic account. In Jordan et al. [2000].

Shaobing Chen et David Donoho. Basis pursuit. In *presented in the 28th Asilomar Conference on Signal, Systems and Computers*, 1994.

C. C. Chevalier. Neural global connectivity in borgesian optimality. *Int. J. of Neural Mar.*, 2000.

P. Comon. Independent component analysis, a new concept ? *Signal Processing*, 36(3) :287–314, 1994.

John Daugman et Cathryn Downing. Gabor wavelets for statistical pattern recognition. Arbib [1995], pages 414–9.

Geoffrey Davis. *Adaptive Nonlinear Approximations.* PhD thesis, 1994.

Rob R. de Ruyter van Steveninck, Geoffrey D. Lewen, Steven P. Strong, Roland Koberle, et William Bialek. Reproducibility and variability in neural spike trains. *Science*, 275 :1805–1808, 1997.

D. Debanne, D. Shulz, et Y. Fregnac. Temporal constraints in associative synaptic plasticity in hippocampus and neocortex. *Can. J. Physiol. and Pharmacol*, 73 :1295–1311, 1995.

S. Deiss, R. Douglas, et A. Whatley. A pulse-coded communications infrastucture for neuromorphic systems. In Maass et Bishop [1999], chapter 6, pages 159–77. URL `http://www.cis.tu-graz.ac.at/igi/maass/PNN.html`.

Arnaud Delorme. *Traitement visuel rapide de scènes naturelles chez le singe, l'homme et la machine : une vision qui va de l'avant.* PhD thesis, Université Paul Sabatier, Toulouse (France), 2001.

Arnaud Delorme, J. Gautrais, R. Van Rullen, et Simon J. Thorpe. Spikenet : a simulator for modeling large networks of integrate and fire neurons. *Neurocomputing*, 24 :663–70, 1999a.

Arnaud Delorme, S. Makeig, Fabre-Thorpe, et Terrence J. Sejnowski. From single-trials eeg to brain area dynamics. *submitted*, 2001a.

Arnaud Delorme, Laurent Perrinet, Simon J. Thorpe, et Manuel Samuelides. Network of integrate-and-fire neurons using rank order coding B : spike timing dependant plasticity and emergence of orientation selectivity. *Neurocomputing*, 38-40(1-4) :539–45, 2001b. URL `http://laurent.perrinet.free.fr/publi/delorme01.pdf`.

Arnaud Delorme, Guillaume Richard, et Michèle Fabre-Thorpe. Rapid processing of complex natural scenes : a role for the magnocellular pathway. *Neurocomputing*, 26-7 :663–70, 1999b.

Arnaud Delorme, Guillaume Richard, et Michèle Fabre-Thorpe. Ultra-rapid categorization of natural scenes does not rely on colour cues : A study in monkeys and humans. *Vision Research*, 40(16) :2187–200, 2000.

Arnaud Delorme et Simon J. Thorpe. Early cortical orientation selectivity : How fast shunting inhibition decodes the order of spike latencies. *submitted*, 2001a.

Arnaud Delorme et Simon J. Thorpe. Event-driven simulation of large networks of spiking neurons. *submitted*, 2001b.

Arnaud Delorme et Simon J. Thorpe. Face processing using one spike per neuron : resistance to image degradation. *Neural Networks*, 6-7(14) :795–804, 2001c.

R. Deriche. Using canny's criteria to derive a recusively implemented optimal edge detector. *Int. J. Computer Vision*, pages 167–87, 1987.

Alain Destexhe. Conductance-based integrate-and-fire models. *Neural Computation*, 9(3) :503–14, 1997.

C. Enroth-Cugell et J. G. Robson. The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, (187) :517–23, 1966.

Michèle Fabre-Thorpe, Arnaud Delorme, Catherine Marlot, et Simon J. Thorpe. A limit in the speed of processing in ultra-rapid categorization of novel natural scenes. *J. Cog. Neurosci.*, 13(2), February 2001.

Daniel E. Feldman. Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron*, 27 :45–56, 2000.

David J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Optical Society of America A*, 4(12) : 2379–94, 1987.

David J. Field. What is the goal of sensory coding ? *Neural Computation*, 6 (4) :559–601, 1994.

P. Foldiak et M. Young. Sparse coding in the primate cortex. Arbib [1995], pages 895–8.

W. T. Freeman, E. H. Adelson, et E. P. Simoncelli. Steerable filters for image analysis. In *Optical Society of America, Annual Meeting*, volume 18, Orlando, October 1989.

P. Fries, S. Neuenschwander, A. K. Engel, R. Goebel, et W. Singer. Rapid feature selective neuronal synchronization through correlated latency shifting. *Nature Neuroscience*, 4 :194–200, 2001.

P. Fries, J. H. Schroder, P. R. Roelfsema, W. Singer, et A. K. Engel. Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. *The Journal of Neuroscience*, 22(9) :3739–54, May 2002.

Dennis Gabor. Theory of communication. *J. of the Institute of Electrical Engineers*, 93(III) :429–57, 1946.

J. L. Gallant, C. E. Connor, S. Rakshit, Lewis J. W., et van Essen D.C. Neural response to polar hyperbolic and cartesian gratings in area v4 of the macaque monkey. *The Journal of Neuroscience*, 76 :2718–39, 1996.

Jacques Gautrais. *Théories et Simulations d'un nouveau type de codage impulsionnel pour le traitement visuel rapide : le codage par l'ordre d'activation.* Sciences cognitives, spécialité : neurosciences computationnelles, École des Hautes Études en Sciences Sociales, 1997.

Jacques Gautrais et Simon J. Thorpe. Rate coding vs. temporal order coding. *Biosystems*, 1997.

Wulfram Gerstner, Richard Kempter, J. Leo Van Hemmen, et Hermann Wagner. Spiking neurons. In Maass et Bishop [1999], chapter 1, pages 3–54. URL `http://www.cis.tu-graz.ac.at/igi/maass/PNN.html`.

Wulfram Gerstner, Richard Kempter, et Wagner Van Hemmen. A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 384 :76–8, 1996.

Simon Haykin. *Neural Networks : A Comprehensive Foundation.* Macmillan College Publishing, New York, second edition, 1999.

Donald O. Hebb. *The Organization of Behavior.* Wiley & Sons, New York, 1949.

J. K. Hietanen, D.I Perret, M.W. Oram, P.J. Benson, et W.H. Dittrich. The effects of lighting conditions on responses of cells selective fo face views in the macaque temporal cortex. *Exp. Brain Res.*, 89 :157–71, 1992.

M. L. Hines et N. T. Carnevale. The NEURON simulation environment. *Neural Computation*, 9(6) :1179–209, 1997. URL `http://neuron.duke.edu`.

Alan Lloyd Hodgkin et Andrew Fielding Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (London)*, 117 :500–44., 1952.

J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci.*, 79 :2554–8, 1982.

David Horn et Sharon Levanda. Fast temporal encoding and decoding with spiking neurons. *Neural Computation*, 10(7) :1705–20, 1999.

Patrik O. Hoyer et Aapo Hyvãrinen. A multilayer sparse coding network learns contour coding from natural images. *To appear in Vision Research*, 2002.

David Hubel et Torsten Wiesel. Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148 :574–91, 1959.

David Hubel et Torsten Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160 :106–54, 1962.

Michel Imbert. La neurobiologie de l'image. *La recherche*, 14 :600–13, 1983.

Giacomo Indiveri. A current-mode hysteretic winner-take-all network, with excitatory and inhibitory coupling. *Analog Integrated Circuits and Signal Processing*, 28 :279—-91, 2001. URL `http://citeseer.nj.nec.com/indiveri01currentmode.html`.

Laurent Itti, Christof Koch, et Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11) :1254–1259, 1998. URL `http://citeseer.nj.nec.com/itti98model.html`.

S. Jaggi et al. High resolution pursuit for feature extraction. Technical memo, Massachusetts Institute of Technology, 1995.

D. A. Jeffreys. A fac-responsive potential recorded from the human scalp. *Exp. Brain Res*, 78(1) :193–202, 1989.

Michael I. Jordan, Michael J. Kearns, et Sara A. Solla, editors. *Advances in Neural Information Processing Systems*, volume 12, 2000. The MIT Press, Cambridge, MA.

Gaetano Kanizsa, editor. *Organization in Vision.* Praeger, New York, 1979.

Richard Kempter, Wulfram Gerstner, et J. Leo Van Hemmen. Hebbian learning and spiking neurons. *Phys. Rev.*, E 59 :4498–4514., 1999.

Richard Kempter, Wulfram Gerstner, et J. Leo Van Hemmen. Intrinsic stabilization of output rates by spike-based hebbian learning. *Submitted to Neural Comput. Dec. 1999, revised Sep. 2000.*, 2000. (1)definition of hebbian learning, refs on stdhp (2)possible scenarios (3)rate based learning (4)spike based.

C. Keysers, D. Xiao, P. Foldiak, et D.I. Perret. The speed of sight. *J. Cog. Neurosci.*, 13(1) :90–101, 2000.

Christof Koch, editor. *Biophysics of Computation : Information Processing in Single Neurons.* Oxford University Press, New York, 1998.

Christof Koch et Idan Segev, editors. *Methods in Neuronal Modeling : From Synapses to Networks.* The MIT Press, Cambridge, MA, Cambridge, Massachusetts, second edition, 1999.

Christof Koch et Idan Segev. The role of single neurons in information processing. *Nature Neuroscience*, 3 :1171–7, 2000.

T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43 :59–69, 1982.

G. Kreiman, I. Fried, et Christof Koch. Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proc. Natl. Acad. Sci. USA*, 99 :8378–83, 2002.

L. Lapicque. Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *J. Physiol. (Paris)*, 9 :620–35, 1907.

S. B. Laughlin. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.*, 36c :910–2, 1981.

D.A. Leopold, M. Wilke, A. Maier, et N.K. Logothetis. Stable perception of visually ambiguous patterns. *Nature Neuroscience*, 5 :605–9, 2002.

AG Leventhal, Y-C Wang, MT Schmolesky, et Y Zhou. Neural correlates of boundary perception. *Vis. Neuroscience*, 15 :1107–1118., 1998.

Zhaoping Li. A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10(4) :903–940, 1998. URL `http://citeseer.nj.nec.com/li98neural.html`.

Ralph Linsker. From basic network principles to neural architecture : Emergence of spacial-opponent cells / orientation-selective cells / orientation columns. *Proceedings of the National Academy of Sciences*, 83 :7508–7512, 8390–8394, 8779–8783, 1986.

Nikos K. Logothetis, D. A. Leopold, et D. L Sheinberg. What is rivalling during binocular rivalry ? *Nature*, 380(6575) :621–4, 2001a.

Nikos K. Logothetis, Jon Pauls, Mark Augath, Torsten Trinath, et Axel Oeltermann. Neurophysiological investigation of the basis of the fmri signal. *Nature*, 412 :150–7, July 2001b.

Wolfgang Maass et Christopher M. Bishop, editors. *Pulsed Neural Networks.* The MIT Press, Cambridge, MA, 1999. URL `http://www.cis.tu-graz.ac.at/igi/maass/PNN.html`.

Wolfgang Maass et A. Zador. Computing with spiking neurons. In Maass et Bishop [1999], pages 55–86. URL `http://www.cis.tu-graz.ac.at/igi/maass/PNN.html`.

K. MacLeid, A. Bäcker, et G. Laurent. Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature*, 390 :70–4, 1998.

Prasanta Chandra Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India.*, 12 :49–55, 1936.

S.F. Mainen et Terrence J. Sejnowski. Reliability of spike timing in neocortical neurons. *Science*, 268 :1503—-6, 1996.

Scott Makeig. Tonic, phasic, and transient eeg correlates of auditory awareness in drowsiness. *Cognitive Brain Research*, 4 :15–25, 1996. URL `http://citeseer.nj.nec.com/28343.html`.

S. Mallat et S. Zhong. Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(7) : 710–732, 1992a.

Stéphane Mallat. *A Wavelet Tour of signal Processing.* Academic Press, 1998.

Stéphane Mallat et Wen Liang Hwang. Singularity detection and processing with wavelets. Technical report, Courant Institute of Mathematical Sciences, New York University, New York, 1991.

Stéphane Mallat et Zhifeng Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12) :3397–3414, 1993.

Stéphane Mallat et Sifen Zhong. Wavelet transform maxima and multiscale edge, 1992b.

H. Markram, J. Luebke, M. Frotscher, et B. Sakmann. Regulation of synaptic efficacy by concidence of postsynaptic APs and EPSPs. *Science*, 275 :213–215, 1997.

H. Markram et M. Tsodyks. Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382 :807–810, 1996.

D. Marr. Visual information processing : The structure and creation of visual representations. *Phil. Trans. R. Soc. Lond., B.*, 290 :199–218, 1980.

D. Marr et E. Hildreth. Theory of edge detection. *Proc. R. Soc. London, Bulletin*, 207 :187–217, 1980.

David Marr. *Vision.* W. H. Freeman and Company, NY, 1982.

J. L. McClelland, David E. Rumelhart, et the PDP Research Group. *Parallel Distributed Processing : Explorations in the Microstructure of Cognition.*, volume 2. The MIT Press, Cambridge, MA, 1986.

W. S. McCullough et W. Pitts. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5 :115–133, 1943.

Carver Mead. *Analog VLSI and Neural Systems.* Addison Wesley, Reading, 1989.

Markus Meister et Michael J. Berry II. The neural code of the retina. *Neuron*, 22 :435—-50, 1999.

Markus Meister et Toshihiko Hosoya. Are retinal ganglion cells independent encoders ? *Nature*, 2001.

C. Meunier et J.-P. Nadal. Sparsely coded neural networks. In *The handbook of Brain Theory and Neural Networks* Arbib [1995], pages 899–901.

Stéphane Mélis. Codage impulsionnels par rangs. Technical report, Rapport de Fin d'Étude, SUPAERO, Toulouse, France, 1995.

C. Monier, L. Borg, et Y. Fregnac. The dominant increase in synaptic conductance evoked by visual input in cat area 17 is due to shunting inhibition. 1997.

Lionel G. Nowak et Jean Bullier. The timing of information transfer in the visual system. In Plenum Press J. Kaas et al., New York, editor, *Extrastriate cortex in primates*, pages 205–41. 1997.

Bruno Olshausen et David J. Field. Natural image statistics and efficient coding. *Network*, 7 :333–339, 1996. URL `http://citeseer.nj.nec.com/olshausen96natural.html`.

Bruno Olshausen et David J. Field. Sparse coding with an overcomplete basis set : A strategy employed by V1 ? *Vision Research*, 37 :3311–25, 1998.

Bruno Olshausen, Phil Sallee, et Michael S. Lewicki. Learning sparse wavelet codes for natural images. In Jordan et al. [2000].

J. Kevin O'Regan et Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(5), 2001.

Stefano Panzeri, Alessandro Treves, Simon Schultz, et Edmund T. Rolls. On decoding the responses of a population of neurons from short time windows. *Neural Computation*, 11(7) :1553–1577, 1999.

Y. Pati, R. Rezaiifar, et P. Krishnaprasad. Orthogonal matching pursuit : Recursive function approximation with applications to wavelet decomposition. In *Proceedings of the 27 th Annual Asilomar Conference on Signals, Systems, and Computers*, 1993. URL `http://citeseer.nj.nec.com/pati93orthogonal.html`.

Laurent Perrinet. Apprentissage hebbien d'un réseau de neurones asynchrone à codage par rang. Technical report, Rapport de stage du DEA de Sciences Cognitives, CERT, Toulouse, France, 1999.

Laurent Perrinet, Arnaud Delorme, Simon J. Thorpe, et Manuel Samuelides. Network of integrate-and-fire neurons using rank order coding A : how to implement spike timing dependant plasticity. *Neurocomputing*, 38–40 (1–4) :817–22, 2001. URL `http://laurent.perrinet.free.fr/publi/perrinet01.pdf`.

Laurent Perrinet et Manuel Samuelides. A generative model for spike time dependent hebbian plasticity. In *Proceedings of DYNN, Bielefeld*, 2000. URL `http://laurent.perrinet.free.fr/publi/perrinet00dynn.pdf`.

Laurent Perrinet et Manuel Samuelides. Coherence detection in a spiking neuron via hebbian learning. *Neurocomputing*, 44–6(C) :133–9, 2002a. URL `http://laurent.perrinet.free.fr/publi/perrinet02.pdf`.

Laurent Perrinet et Manuel Samuelides. Sparse image coding using an asynchronous spiking neural network. In *Proceedings of ESANN*, pages 313–8, 2002b. URL `http://laurent.perrinet.free.fr/publi/perrinet02esann.pdf`.

Laurent Perrinet et Manuel Samuelides. Visual strategies for sparse spike coding. In *Actes de Neurosciences et Sciences de l'Ingénieur, L'Agelonde.*, 2002c. URL `http://laurent.perrinet.free.fr/publi/perrinet02nsi.pdf`.

Laurent Perrinet, Manuel Samuelides, et Simon Thorpe. Sparse spike coding in an asynchronous feed-forward multi-layer neural network using matching pursuit. *Neurocomputing*, 2002. URL `http://laurent.perrinet.free.fr/publi/perrinet02sparse.pdf`.

E. Peterhans et R. Von der Heyt. Mechanisms of contour perception in monkey visual cortex. II contours bridging gaps. *The Journal of Neuroscience*, 9 :1749–1763, 1989.

C.V. Portfors et J.J. Wenstrup. Delay-tuned neurons in the inferior colliculus of the mustached bat : implications for analyses of target distance. *Journal of Neurophysiology*, 82 :1326–38, 1999.

Charles Poynton. Frequently asked questions about gamma. Technical report, 1999. URL `http://www.inforamp.net/~poynton`.

Alexander D. Protopapas, Michael Vanier, et James M. Bower. Simulating large networks of neurons. In Koch et Segev [1999], chapter 12, pages 461–498.

John Rinzel et Bard Ermentrout. Analysis of neural excitability and oscillations. In Koch et Segev [1999], chapter 7, pages 251–292.

O. Rochel, D. Martinez, E. Hugues, et F. Sarry. Stereo-olfaction with a sniffing neuromorphic robot using spiking neurons. In *Eurosensors*, 2002.

R. W. Rodieck. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5 :583–601, 1965.

E. Rodriguez, N. George, J.-P. Lachaux, J. Martinerie, B. Renault, et F. Varela. Perception's shadow : long-distance gamma band synchronizationof human brain activity. *Nature*, 397 :430–3, 1999.

Rosenblatt. Perceptron simulation experiments. *Proceedings of the I. R. E.*, 20 :167–192, 1960.

Botond Roska et Frank Werblin. Vertical interactions across ten parallel, stacked respresentations in the mammalian retina. *Nature*, 410 :583–7, 2001.

Mark C. W. Van Rossum, G. Q. Bi, et G. G. Turrigiano. Stable hebbian learning from spike timing-dependent plasticity. *The Journal of Neuroscience*, 20(23) :8812–21, 2000.

G.A. Rousselet, M. Fabre-Thorpe, et S.J. Thorpe. Two unrelated natural scenes can be processed as fast as one. *ECVP 2001. Supplement.*, 2002.

J. Rubin, D. Lee, et H. Sompolinsky. Equilibrum properties of temporally asymmetric hebbian plasticity. In *ESF Symposium on Neuronal Plasticity and Dynamics, Trieste.*, 2000.

Emilio Salinas et Terrence J. Sejnowski. Correlated neuronal activity and the flow of neural information. *Nature reviews, Neuroscience*, 2 :539–554, 2001.

Ferdinando Samaria et Andy Harter. Parameterisation of a stochastic model for human face identification. In *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota FL, December 1994*, 1994. URL `http://www.uk.research.att.com/pub/docs/att/paper.95.2.ps.Z`.

Manuel Samuelides, Simon J. Thorpe, et E. Veneau. Implementing Hebbian learning in a rank-based neural network. *Lecture Notes in Computer Science.*, 1327 :145–50, 1997.

Maria V. Sanchez-Vives, Lionel G. Nowak, et David A. McCormick. Membrane mechanisms underlying contrast adaptation in cat area 17 in vivo. *The Journal of Neuroscience*, 10(11) :4267—-85, 2000. URL `http://www.cns.nyu.edu/earclub/artfiles/sanchez-vives2000.pdf`.

Walter Senn, Misha Tsodyks, et Henry Markram. An algorithm for synaptic modification based on exact timing of pre- and post-synaptic action potentials. In *ICANN'97*, pages 121–126, 1997.

C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27 :379–423, 623–56, 1948.

Roger N. Shepard et Jacqueline Metzler. Mental rotation of three-dimensionnal objects. *Science*, 171 :701–4, 1970.

S.M. Sherman et R.W. Guillery. Functional organization of thalamocortical relays. *Journal of Neurophysiology*, 76 :1367–95, 1996.

Mariano Sigman, Guillermo A. Cecchi, Charles D. Gilbert, et Marcelo O. Magnasco. On a common circle : Natural scenes and gestalt rules. *PNAS*, 98(4) :1935–40, February 2001.

E P Simoncelli et W T Freeman. The steerable pyramid : A flexible architecture for multi-scale derivative computation. In *Second Int'l Conf on Image Processing*, Washington, DC, October 1995.

William Softky et Christof Koch. Cortical cells should fire regularly, but do not. *Neural Computation*, 4 :643–646, 1993a.

William Softky et Christof Koch. The highly irregular firing of cortical cells is inconsistence with temporal integration of random epsps. *The Journal of Neuroscience*, 13 :334–350, 1993b.

Sen Song, Kenneth D. Miller, et Larry F. Abbott. Competitive Hebbian learning through spike-timing dependent synpatic plasticity. *Nature Neuroscience*, 3 :919–26, 2000.

Charles F. Stevens et Anthony M. Zador. Input synchrony and the irregular firing of cortical neurons. *Nature Neuroscience*, 1(3), 1998.

G. J. Stuart et B. Sakmann. Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature*, 367(6458) :69–72, 1994.

Ivan Sutherland et Jon Lexau. Designing fast asynchronous circuits. In *Proceedings of the Seventh International Symposium on Advanced Research in Asynchronous Circuits and Systems, Salt Lake City, Utah, USA, March 2001.*, pages 184–93. Copyright ©2001 by IEEE., 2001.

R.S. Sutton et A.G. Barto. Time derivative models of pavlovian reinforcement. In M. Gabriel et J.W. Moore, editors, *Learning and Computational Neuroscience*. The MIT Press, Cambridge, MA, 1991.

K. Tanaka. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.*, 19 :109–39, 1996.

Simon J. Thorpe. Spike arrival times : a highly efficient coding scheme for neural networks. *Parallel processing in Neural Systems and Computers*, pages 91–4, 1990.

Simon J. Thorpe, A. Delorme, et R. VanRullen. Spike based strategies for rapid processing. *Neural Networks*, 6-7(14) :715–26, 2001a.

Simon J. Thorpe, Denis Fize, et Catherine Marlot. Speed of processing in the human visual system. *Nature*, 381 :520–2, 1996.

Simon J. Thorpe et Jacques Gautrais. Rank order coding. In *Computational Neuroscience : Trends in Research 1998, J. Bower, Editor. Plenum Press : New York.*, pages 113–8. 1998.

Simon J. Thorpe, K. R. Gegenfurtner, Michèle Fabre-Thorpe, et H. H. Bulthoff. Detection of animals in natural images using far peripheral vision. *Eur. J. Neuroscience*, 14(5) :869–76, September 2001b.

Simon J. Thorpe et M. Imbert. Biological constraints on connectionist modelling. In *Connectionism in Perspective*. Pfeifer, R., Schreter, Z., Fogelman-Soulié, F., and Steels, L., eds., Elsevier (North-Holland), 1989.

Simon J. Thorpe et Michel Imbert. *Parallel processing in neural systems.* R.Eckmiller G.Hartman and G.Hauske, North Holland, 1990.

T. W. Troyer, A. Krukowski, N. J. Priebe, et K. D. Miller. Contrast-invariant orientation tuning in cat visual cortex : Feedforward tuning and correlation-based intracortical connectivity. *The Journal of Neuroscience*, 18 :5927, 1998.

Misha Tsodyks, Klaus Pawelzik, et Henry Markram. Neural networks with dynamic synapses. *Neural Computation*, 10(4) :821–835, 1998.

Antonio Turiel, Germán Mato, Néstor Parga, et Jean-Pierre Nadal. Self-similarity properties of natural images. In Michael I. Jordan, Michael J. Kearns, et Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. The MIT Press, Cambridge, MA, 1998.

Antonio Turiel et Nestor Parga. The multifractal structure of contrast changes in natural images : from sharp edges to textures. *Neural Computation*, 12 :763–93, 2000.

A. Turing. The chemical basis of morphogenesis. *Phil. Trans. of the Roy. Soc.*, B, 1952.

L. Ungerleider et R. Desimone. Projections to the superior temporal sulcus from the central and peripheral representation of V1 & V2. *Journal of Comp. Neuroscience*, 248 :164–89, 1986.

J. Vaccaro, D. Gourion, M. Samuelides, et S. Thorpe. Rank based hebbian learning in a multi-layered neural network. In *Proccedings of VI-DYNN'98. Royal Institute of Technology, Stockholm, Sweden.*, 1998.

J.H. van Hateren. Spatiotemporal contrast sensitivity of early vision. *Vision Research*, 33 :257–67, 1993.

J.H. van Hateren et H.P. Snippe. Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells. *Vision Research*, 41 :1851–65, 2001.

J.H. van Hateren et A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc.R.Soc.Lond.B*, 265 :359–66, 1998.

Mark C. W. van Rossum, Gina G. Turrigiano, et Sacha B. Nelson. Fast propagation of firing rates through layered networks of noisy neurons. *The Journal of Neuroscience*, 22(5) :1956—-66, 2001.

Rufin Van Rullen. *Une première vague de potentiels d'actio, une première vague idée de la scène visuelle. Rôle de l'asynchronie dans le traitement rapide de l'information visuelle.* PhD thesis, 2001.

Rufin Van Rullen, Arnaud Delorme, et Simon J. Thorpe. Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation. *Neurocomputing*, 1-4(38-40) :1003–9, 2001.

Rufin Van Rullen, Jacques Gautrais, Arnaud Delorme, et Simon J. Thorpe. Face processing using one spike per neuron. *BioSystems*, 483 :229–39, 1998.

Rufin Van Rullen et Simon J. Thorpe. Spatial attention in asynchronous neural networks. *Neurocomputing*, 26–7 :911–8, 1999.

Rufin Van Rullen et Simon J. Thorpe. Is it a bird ? is it a plane ? Ultra-Rapid Visual Categorisation of natural and artifactual objects. *Submitted to Perception*, 2000.

Rufin Van Rullen et Simon J. Thorpe. Rate coding versus temporal order coding : What the retina ganglion cells tell the visual cortex. *Neural Computation*, 13(6) :1255–83, 2001.

Rufin Van Rullen et Simon J. Thorpe. Surfing a spike wave down the ventral stream. *Vision Reseach*, page in press, 2002.

F. Varela, J.-P. Lachaux, E. Rodriguez, et J. Martinerie. The brainweb : phase synchronization and large-scale integration. *Nat. Rev. Neurosci.*, 2 (4) :229–39, 2001.

Emmanuel Veneau. Codage impulsionnels par rangs et apprentissage. Technical report, Rapport de Fin d'Étude, SUPAERO, Toulouse, France, 1996.

William E. Vinje et Jack L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287 :1273–1276, 2000.

R. Vogels. Categorization of complex visual images by rhesus monkeys. *Eur. J. Neurosci.*, 11(4) :1223–55, 1999.

G. Wallis et E. T. Rolls. Invariant face and object recognition in the visual system. *Prog. Neurobiology*, 51 :167–194, 1997.

Xiaohui Xie et Sebastian H. Seung. Spike-based learning rules and stabilization of persistent neural activity. In Jordan et al. [2000].

A. Zador. Impact of synaptic unreliability on the information transmitted by spiking neurons. *J. Neurophysiol.*, 79 :1219–29, 1998. URL `http://citeseer.nj.nec.com/zador98impact.html`.

S. Zeki et S. Shipp. The functional logic of cortcal connections. *Nature*, 335 : 311–7, 1998.

Résumé : ***Comment déchiffrer le code impulsionnel de la Vision ? Étude du flux parallèle, asynchrone et épars dans le traitement visuel ultra-rapide.***

Le cadre de ce travail est l'étude de modèles neuromimétiques de codage parallèle et asynchrone de l'information visuelle —tel qu'il est mis en évidence dans des taches de traitement ultra-rapide— en la transformant en une vague d'événements élémentaires d'importance décroissante. Nous allons baser dans un premier temps les mécanismes de ce code sur les processus biologiques à l'échelle du neurone et de la synapse. En particulier, la plasticité synaptique peut induire l'extraction non-supervisée de l'information cohérente dans le flux des impulsions neuronales.

Le codage par la latence de la première décharge permet de définir un code impulsionnel dans le nerf optique grâce une architecture multiéchelle. Nous avons étendu cette démarche en utilisant une approche *écologique* qui permet exploiter les régularités de ses coefficients sur les images naturelles pour les quantifier par le rang d'arrivée des impulsions neuronales. Ce code par le rang des décharges, est basé sur une architecture hiérarchique et "en avant" qui se distingue, outre sa simplicité, par la richesse des résultats mathématiques et de par ses performances computationnelles.

Enfin, nous avons répondu aux besoins d'un modèle efficace de la Vision en fondant une théorie de *représentation impulsionnelle sur-complète* de l'image. Cette formalisation conduit alors à une stratégie de *code impulsionnel épars* en définissant des interactions latérales. Cette stratégie est étendue à un modèle général de *colonne corticale adaptative* permettant l'émergence de dictionnaires de représentation et s'adapte particulièrement à la construction d'une carte de saillance. Ces techniques font émerger de nouveaux outils pour le traitement de l'image et de vision active adaptés à des architectures de calcul distribué.

Mots clés : Vision, codage neuronal impulsionnel, plasticité hebbienne à modulation temporelle relative (STDP), transformée par ondelettes, statistique des images naturelles, représentation sur-complète, poursuite de concordance, émergence, vision active.

Abstract : ***How to decipher vision's spiking code ? Study of the parallel, asynchronous and sparse flow in the ultra-rapid visual processing.***

We build and study dynamical models of visual coding as a parallel and asynchronous flow of information coded thanks to their succession in time. We will at first base the mechanisms of this code on the biological processes on the scale of the neurone and synapse. In particular, synaptic plasticity may induce the non-supervised extraction of coherent information in the flow of the neuronal impulses.

Coding by the latency of the first spike can defines a code in the optic nerve based on multi-scale architecture. We extended these results by using an ecological approach allowing thanks to the *statistics of natural images* the quantization of analog value by the spikes' rank. This visual code is based on a hierarchical feed-forward architecture which is distinguished, in addition to its simplicity, by its mathematical and computational performances.

We will meet the needs for an effective model of Vision by defining a theory of *overcomplete event representation* of the image. This formalization leads then to a strategy of a *sparse spike code* by defining lateral interactions. Thanks to a reinforcement learning rule, this strategy can then be extended to a model of an *adaptive cortical column* which shows emergence of representation dictionaries. Moreover, this paradigm adapts particularly to the construction of a saliency map. These techniques allow emergence of new tools for image processing and active vision which are particularly adapted to distributed computing architectures.

Keywords : Vision, parallel and asynchronous computing, neuronal spike coding, spike time dependent plasticity (STDP), wavelet transform, natural images statistics, overcomplete representation, matching pursuit, active vision.